GSBPM

Version 1.0

This document includes guidelines aiming to help the staff of CYSTAT when completing the questionnaire regarding the documentation of the various processes followed during all the phases of statistical production, in accordance with the Generic Statistical Business Process Model (GSBPM).

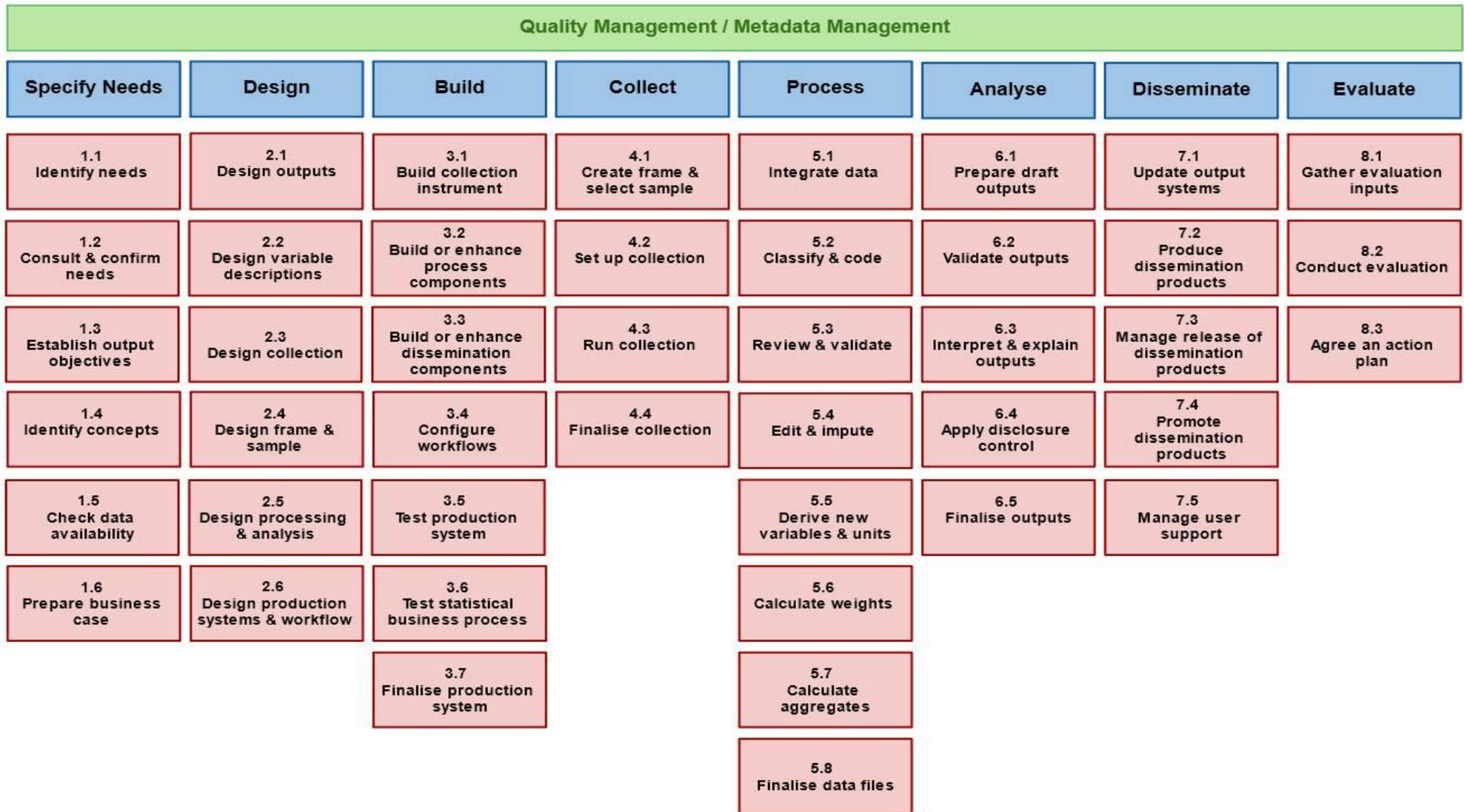*GUIDELINES FOR IMPLEMENTATION AT CYSTAT*

Table of Contents

I.   INTRODUCTION

1.   The Generic Statistical Business Process Model (GSBPM) describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonised terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components.

2.   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, as well as to share methods and components.

3.   Following the example of many other NSIs who have successfully implemented GSBPM and have based their quality management system on this, CYSTAT has decided to also use GSBPM in order to record the processes followed for statistical production. The documentation of processes in CYSTAT (and within any organisation for that matter) is expected to yield many benefits, such as:
- Ensuring business continuity;
- Clarity and transparency regarding the processes followed in statistical production;
- Recording useful information to be used when developing the Data-Warehouse of CYSTAT;
- Achieving standardised and harmonised procedures through the analysis and improvement of current procedures;
- Issuing quality guidelines regarding statistical processes;
- Improving the overall efficiency of the statistical production.

4.   The GSBPM comprises three levels:
- Level 0, the statistical business process;
- Level 1, the eight phases of the statistical business process;
- Level 2, the sub-processes within each phase.

5.   A diagram showing the phases (level 1) and sub-processes (level 2) of the GSBPM is presented in the next page (Figure 1). The sub-processes are described in detail in Section II.

6.   Due to the significant amount of time needed to provide descriptions for all sub-processes defined, the collection of information within CYSTAT will take place in two stages. At a first stage, the production units will be asked to provide information on selected sub-processes deemed to be of primary importance. Information on the remaining sub-processes will be provided by the production units at a later stage.

7.   The official GSBPM manual (version 5.0) is available at CYSTAT's SharePoint site at the following link: https://govcloud.gov.cy/ministries/MOF/cystat/quality/GSBPM. The current document with CYSTAT's specific guidelines for completing the national GSBPM questionnaire is also stored at the same web-address.

8.   For more information regarding the GSBPM, one may visit UNECE's website at the following link: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0.

Quality Management / Metadata Management

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

II. CYSTAT GSBPM QUESTIONNAIRE AND GUIDELINES

The structure of the questionnaire is such that the official definition for each phase and sub-process is provided, followed by a set of questions aiming to cover all the relevant information. As mentioned in Section I, at a first stage, only a selected number of questions deemed of highest priority need to be answered. The remaining questions (disabled in the questionnaire at this stage) will be answered at a later stage.

GENERAL INFORMATION

G.1   Insert the name of the survey/statistical activity (as reported in the annual programme of work). Do not include the reference period in the title:

G.2   Insert the reference period for the survey/statistical activity (year, quarter, month etc.):

G.3   Code of survey/statistical activity (to be inserted by the working group):
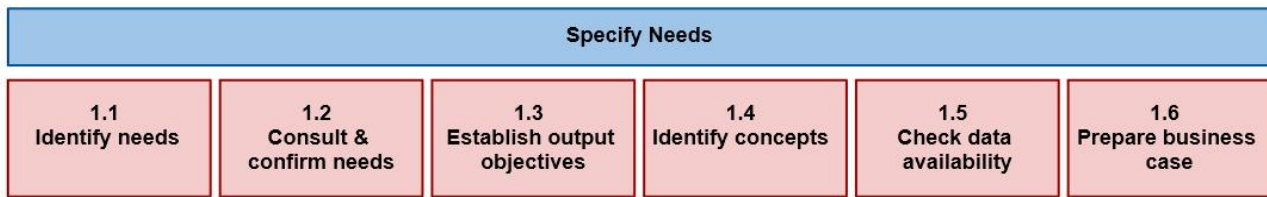

G.4   Date of submission of this questionnaire (dd/mm/yyyy):

G.5   Name of the person completing the questionnaire:

G.6   Name of the responsible officer(s) for the specific survey/statistical activity:

G.7   CYSTAT Division/Section responsible for the specific survey/statistical activity:

PHASE 1 – SPECIFY NEEDS

Specify Needs					
1.1 Identify needs	1.2 Consult & confirm needs	1.3 Establish output objectives	1.4 Identify concepts	1.5 Check data availability	1.6 Prepare business case

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

In this phase the organisation:
- identifies the need for the statistics;
- confirms, in more detail, the statistical needs of the stakeholders;
- establishes the high level objectives of the statistical outputs;
- identifies the relevant concepts and variables for which data are required;
- checks the extent to which current data sources can meet these needs;
- prepares the business case to get approval to produce the statistics.

SUB-PROCESS 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, 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 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.

1.1.1   Scope of the survey – What statistics are needed? Provide the general concepts not specific variables and definitions:

1.1.2   What purpose do these statistics serve?

1.1.3   Relevant EU Regulations:

1.1.4   Relevant National Laws (Other than the Statistics Law No. 15(I) of 2000):

- 7 -

1.1.5   National needs (e.g. Ministry of Finance, Trade Unions, University Research Centres etc.):

1.1.6   International requests (e.g. IMF, Troika etc.):

1.1.7   Internal needs (within CYSTAT):

1.1.8   User community's needs (e.g. disabled, ethnic groups etc.):

1.1.9   Other (Please specify):

1.1.10   To what extent have stakeholders been identified and included in discussions about statistical needs underpinned by this domain?

SUB-PROCESS 1.2: 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.

1.2.1   In relevance to your answer in sub-process 1.1 above regarding national, international, internal user communities needs or any other needs, please provide information on:

a)   How the initial request/user need was made known to CYSTAT (e.g. letters, meetings etc.):

b)   What the procedure for following up on the request was:

1.2.2   Have needs changed after consultation sessions? Provide a description to explain why certain needs were taken on board as opposed to others which might have not:

SUB-PROCESS 1.3: 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.

1.3.1   Outputs agreed in order to meet user needs (e.g. publications, news releases, tables etc.).There is no need to list all tables that will be produced, but the main aspects and thematic areas that will be tackled. For example, 'Unemployment rates by demographics' or 'Earnings by employee demographics' etc.:

1.3.2   Outputs stipulated by EU Regulations, Acts etc.:

1.3.3   Foreseen confidentiality issues in relevance with the agreed outputs (e.g. data relevant to armed forces etc.):

1.3.4   Agreed quality measures relevant to the outputs (e.g. standard errors, response rates etc.)

1.3.5   Is there a possibility for users to apply for tailor-made requests? If no, please specify the reasons:

SUB-PROCESS 1.4: 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. This alignment, and the choice or definition of the statistical concepts and variables to be used, takes place in sub-process 2.2.

1.4.1   Please provide the main concepts requested by users, with brief descriptions:

1.4.2   Level of compliance with existing variables. In case of discrepancies please explain:

SUB-PROCESS 1.5: 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.

1.5.1   Do current data sources (e.g. administrative sources, secondary data etc.) meet the user requirements?

> **1.5.1.1 If no, are you aware of other data sources (administrative of secondary data) that could satisfy user requirements? Please specify:**

> **1.5.1.2 If yes, are these sources already available within CYSTAT?**

> **1.5.1.3 If not available within CYSTAT, are there any restrictions for acquiring the data? Please specify:**

1.5.2   Are there any planned/proposed actions to ensure that new data sources are made available? (E.g. Memorandum of Understanding, provisions to be included in new Statistics Law etc.)

1.5.3   Refer to any secondary data that may be used for benchmarking and checking purposes:

1.5.4   Please specify whether there are current data sources that may be deemed <u>not</u> suitable in the future:

SUB-PROCESS 1.6: PREPARE 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.

1.6.1   Describe 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:

1.6.2   Describe the proposed "To-Be" solution, detailing how the statistical business process will be developed to produce the new or revised statistics:

1.6.3   Analysis of the costs relevant to this business process [insert table]:

1.6.4   Information regarding human resources:

1.6.5   Constraints (budget restrictions, HR restrictions, etc.)

PHASE 2 – DESIGN

Design					
2.1 Design outputs	2.2 Design variable descriptions	2.3 Design collection	2.4 Design frame & sample	2.5 Design processing & analysis	2.6 Design production systems & 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 (any tool or routine to gather or extract data and metadata, from paper questionnaires to web-scraping tools), 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. Organisations are also encouraged to reuse or adapt design elements from existing processes. Additionally, outputs of design processes may form the basis for future standards at the organisation, national or international levels.

SUB-PROCESS 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. 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 from sub-process 1.1 (Identify needs).

2.1.1 What type of dissemination products will be produced? Please select from the following (multiple selection):

☐ News releases
☐ Statistical tables (key figures on CYSTAT website)
☐ Electronic publications (CYSTAT website)
☐ Printed publications
☐ Tailor made data
☐ Methodological information (quality reports, metadata reports)
☐ Micro-data (e.g. transmission to Eurostat, Central Bank of Cyprus etc.)
☐ Macro-data (e.g. transmission to Eurostat, Central Bank of Cyprus etc.)

2.1.2 In the case of News Releases, do you use the predefined CYSTAT template?

2.1.3   In the case of statistical tables please provide the relevant titles:

- 13 -

2.1.4   In the case of a publication (electronic or printed), please provide the relevant title (most recent publication):

2.1.5   In the case of methodological information, which standards are applied? (e.g. ESMS, ESQRS, SDDS etc.)

2.1.6   In the case of micro-data, please specify whether a transmission format exists and provide the title of any relevant documentation:

2.1.7   Are there any standards applied for micro-data? Please specify (e.g. SDMX etc.):

2.1.8   Are there any standards applied for macro-data? Please specify (e.g. SDMX, SDDS etc.):

2.1.9   Describe any disclosure control methods considered during the design phase:

2.1.10 Please describe any preliminary discussions regarding the design of the outputs. with the following sections/units/committees:

a)   Dissemination Section:

b)   Data Processing Unit:

c)   Blaise Team:

d)   Confidentiality Committee:

This sub-process defines the statistical 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 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 instrument may be inter-dependent to some degree. Preparation of metadata descriptions of collected and derived variables and classifications is a necessary precondition for subsequent phases.

2.2.1   Please fill in Annex 2.2 Design Variable Descriptions [variables collected, variables derived, classifications, code lists]

- The purpose is essentially to obtain a list of all variables that are collected and derived (any intermediate calculations that are not stored somewhere are not needed) – anything stored to be used needs to be recorded. Do not report the questions from the questionnaires. Take into consideration administrative sources as well.
- Here, the 'variables' (derived and non-derived) and the corresponding 'definitions' need to be derived (variables not 'concepts'). In this section the FINAL list of variables is to be provided. A distinction must be made between derived and the non-derived variables.
- Refer to any classifications – ISCO, NACE, MGC, ISCED, etc. (one may also provide the link to the respective classification in the NSO Metadata website under 'Classifications').
- Actual lists of variables, code lists, and definitions are to be provided here.

This sub-process determines the most appropriate collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques. This sub-process includes the design of 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 design of any formal agreements relating to data supply, such as memoranda of understanding, and confirmation of the legal basis for the 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).

Note: This sub-process may be revisited after any pilot studies.

2.3.1 What data collection method is planned to be used? (multiple selection):

☐ PAPI (Paper Assisted Personal Interviewing) - Describe how the data entry is planned:

```

```

☐ CAPI (Computer Assisted Personal Interviewing)
☐ CATI (Computer Assisted Telephone Interviewing)
☐ CAWI (Computer Assisted Web Interviewing)
☐ CARI (Computer Assisted Recorded Interviewing)
☐ Postal - Describe how the data entry is planned:

```

```

☐ Email - Describe how the data entry is planned:

```

```

☐ Administrative sources
☐ Results/data of other surveys
☐ Integration of various sources
☐ Other - Please specify:

```

```

2.3.2 Why were the selected methods preferred over other methods?

```

```

2.3.3   Regarding any administrative sources used, how are the data planned to be collected / provided to CYSTAT?

☐ GDW (ΚΑΠ) - DataMart

☐ GDW (ΚΑΠ) - GQ

☐ Email (encrypted)

☐ Email (non-encrypted)

☐ USB

☐ CD/DVD

☐ Direct access to administrative source's information system

☐ Access to administrative source's website

☐ Other (please specify):

2.3.4   Is there a model questionnaire provided (e.g. by Eurostat)? Are there other guidelines supplied (e.g. specification of which questions should be open or closed)?

2.3.5   Questionnaire design: Describe all steps required for the production of the final questionnaire. Also refer to any related templates prepared (such as excel templates to be sent via email).

2.3.6   Are there any deviations from what is stipulated in the relevant legal framework (e.g. European regulations)? Please specify:

Note: Whereas in sub-process 1.5 one should check about the availability and adequacy of the legal framework, in sub-process 2.3 one has to explain what happens in real terms, i.e. any deviations from what is stipulated in the aforementioned EC Regulation(s).

2.3.7   Provide information about any agreements made regarding the provision of administrative data for this data collection (e.g. Memoranda of Understanding, or other agreements e.g. resulting from bilateral meetings or official letters):

2.3.8   Specify any templates designed for the data collection (please attach):
☐Initial letter sent to enterprises/households
☐Letter(s) regarding refusal to cooperate
☐Letters/Notes to be used when respondents cannot be reached
☐Confidentiality agreement (for interviewers) – no attachment needed
☐Contract of employment – no attachment needed
☐Termination of employment
☐Other templates (please specify):

2.3.9   Has an evaluation of the time needed for the respondent to provide the information (response burden) been taken into account when designing the questionnaire?

    2.3.9.1 Please specify:

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. A sampling plan should be made: The actual sample is created in sub-process 4.1 (Create frame and select sample), using the methodology, specified in this sub-process.

2.4.1   Population of interest:

2.4.2   Sampling frame and corresponding reference period (e.g. Business Register – 2016 data up to December 2016).

☐ Business Register
Reference period:

☐ Census of population 2011 / households
Reference period:

☐ Census of population 2011 / individuals
Reference period:

☐ Agricultural Register
Reference period:

☐ Social Insurance Register
Reference period:

☐ Other - please specify:

2.4.3   Are there any differences between the target survey population as specified by European legislation and the actual survey population?

2.4.3.1 Please specify:

2.4.4   Are any administrative data used to supplement sample data?

2.4.4.1 Please specify:

2.4.5   How is the desired sample size for the survey determined? (E.g. based on past data's standard deviation, response rate, population size etc.). Please describe:

2.4.6 Specify the sampling technique to be used for the sample selection (e.g. simple random sampling, probability proportional to size etc.). Also refer to any stratification planned to be applied:

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, estimating, integrating, validating and finalizing data sets.

Describe the methodological procedure to be used for the following:
Provide general information/an overall evaluation of the procedures and methodologies selected. Technical details and detailed descriptions must be supplied at the corresponding sub-processes of Process 5, __not__ here.

Note: Information about disclosure control must not be included here, but in sub-processes 3.3 and 6.4 accordingly.

2.5.1 Coding (Include details about the data entry application. E.g. how is the coding of NACE and ISCO planned to be carried out?):

2.5.2 Editing:

2.5.3 Imputing:

2.5.4 Calculation of weights – adjustments for non-response (basic information):
Note: Technical details about the methods used and how these are calculated in practice must be included in sub-process 5.6.

2.5.5   Calibration techniques (basic information):

2.5.6   Estimating data:

2.5.7   Integrating data sources:

2.5.8   Validating data:

2.5.9   Finalizing data sets (e.g. for publication, transmitting to Eurostat etc.):

2.5.10   Please provide the list of validation checks planned to be used:
    2.5.10.1   During data collection/data entry (include in-built checks):
    ☐ List attached
    ☐ Not applicable

    2.5.10.2   During processing/analysis (e.g. consistency checks, coherence checks etc.):
    ☐ List attached
    ☐ Not applicable

SUB-PROCESS 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 statistical production process, and ensuring that they fit together efficiently with no gaps or redundancies. Various systems and databases are needed throughout the process. 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 process, then, if any gaps are identified, new solutions should be designed. This sub-process also considers how staff will interact with systems, and who will be responsible for what and when.

2.6.1   Describe the sub-processes from data collection to dissemination. Include information about 'who' does 'what' and at 'which stage'.

☐Table attached

Description:

PHASE 3 – BUILD PHASE

Build						
3.1 Build collection instrument	3.2 Build or enhance process components	3.3 Build or enhance dissemination components	3.4 Configure workflows	3.5 Test production system	3.6 Test statistical business process	3.7 Finalise production system

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. These new services are constructed to be broadly reusable within the statistical production architecture.

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.

SUB-PROCESS 3.1: BUILD 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 testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to the statistical metadata system, so that metadata can be more easily captured in the collection phase. Connection of 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.

3.1.1 For the paper format of the questionnaires used, describe the procedure followed (who designs the layout of questionnaire, who creates the questionnaire, if there is a template used, approval procedure, printing pre-requisites if applicable, printing arrangements, dissemination on webpage, format etc.):

3.1.2 For electronic questionnaires (CAPI, CATI, CAWI, Pdf Forms, Excel files etc.) describe the development process. Mention the use of any coding applications (e.g. dictionaries):

3.1.3   In the case of administrative sources describe the procedure followed to obtain the data, if applicable (include references to the software used). Please note that this refers to the procedure of obtaining the data and not the use of the administrative data:

3.1.4   Describe the procedure for pilot-testing, if applicable, and how the results are taken into account:

3.1.5   Does the collection instrument facilitate the estimation of interview time?

3.1.5.1 If "YES", please specify:

SUB-PROCESS 3.2: BUILD OR ENHANCE PROCESS COMPONENTS

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

3.2.1   In the case of Paper Assisted Personal Interviewing (PAPI) or Paper Assisted Telephone Interviewing (PATI) please provide information on the following:

3.2.1.1 IT applications developed for data entry:

3.2.1.2 IT applications developed for coding (<u>not</u> integrated in the data collection phase):

SUB-PROCESS 3.3: 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 as designed in sub-process 2.1 (Design outputs). 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.

Provide information on the process of preparing/building the tools/components/services needed for the following planned (according to sub-process 2.1) dissemination products (e.g. cover design, printing, publication layout, etc.):

3.3.1   Electronic publications:

3.3.2   Printed publications (e.g. printing outsourced, external designer designs front cover, why external help is needed etc.):

3.3.3   News releases:

3.3.4   Statistical tables (key figures on CYSTAT website):

3.3.5   Tailor made data:

3.3.6 Methodological information (quality reports ESQRS, metadata reports ESMS, SDDS):

3.3.7 Micro-data (if SDMX files are produced provide data structure definitions used etc.) For disclosure control of microdata, the corresponding identification levels should be included:

3.3.8 Macro-data (if SDMX files are produced, data structure definitions used etc.):

SUB-PROCESS 3.4: CONFIGURE WORKFLOWS

This sub-process configures the workflow, systems and transformations used within the statistical business processes, from data collection through to dissemination. It ensures that the workflow specified in sub-process 2.6 (Design production systems and workflow) works in practice.

3.4.1 Discuss about the flow from one sub-process to another, as specified in sub-process 2.6 (from data collection to dissemination) to make sure that each sub-process is achievable, even in terms of the timeframes set in sub-process 2.6. Therefore here, one should comment about 'timeliness' i.e. the likelihood that each sub-process is completed in a timely manner:

3.4.2 Provide information about making sure that all required facilities are in place, e.g. recruitment of staff, installation of IT programs, availability of netbooks/laptops/tablets, etc.

SUB-PROCESS 3.5: TEST PRODUCTION SYSTEM

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 3.2 (Build or enhance process components), this sub-process also includes testing of interactions between assembled and configured services, and ensuring that the production solution works as a coherent set processes, information and services.

3.5.1   Provide information with regard to the testing of applications referred to in sub-process 3.2; including checks on IT systems and checks to ensure interaction between all activities (e.g. data entry and coding), as well as interaction with external systems (e.g. primary and secondary data sources) which will both be used:

3.5.2   Provide information on checks on hardware (e.g. check that all laptops are in working order and that any networking is functioning well):

SUB-PROCESS 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 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.

Note: Details of any improvements made should be included in the relevant sub-processes.

3.6.1   Describe how the business process was piloted from start to finish (i.e. through any pilot studies). If a pilot study has been carried out to test specific processes (questionnaire, data collection, data entry, etc.), this should <u>not</u> be included here, but in sub-process 3.1. Here we are interested in any piloting of the whole 'process':

3.6.2   If applicable, provide information on the assessment of major error sources from the pilot (e.g. coverage, non-response, measurement and process errors):

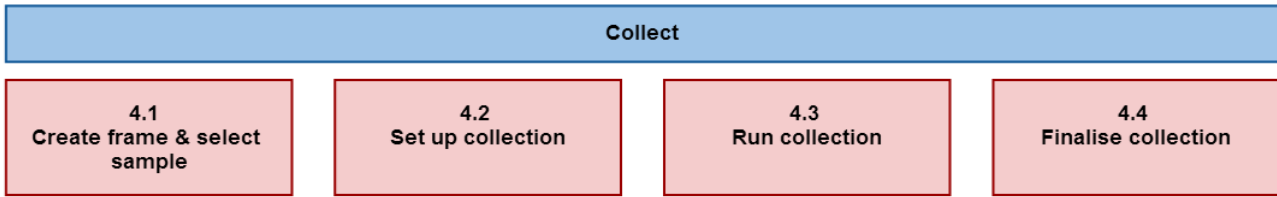**SUB-PROCESS 3.7: FINALISE PRODUCTION SYSTEMS**

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 (this activity may also be part of sub-process 3.5 (Test production system)).

3.7.1   Provide information on any technical manuals/documentation regarding the various tools to be used and any manuals prepared for business users:

3.7.2   Provide information regarding any training for business users (not interviewers involved in data collection), for example training of survey managers in the use/import/export procedures to be followed with BLAISE, or other applications developed:

PHASE 4 – COLLECT

Collect			
4.1 **Create frame & select sample**	**4.2** **Set up collection**	**4.3** **Run 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. 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.

SUB-PROCESS 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 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) as such processes generally create frames from the available data and then follow a census approach.

Note: Here, the sample methodology specified in sub-process 2.4 is confirmed, and the focus is on the actual selection process.

4.1.1   Please provide a description of creating the <u>frame</u> for the specific statistical activity (irrelevant in the cases were no sample is drawn or for domains based entirely on administrative data). Confirm that the steps defined in sub-process 2.4 were followed. If not, please describe any differences from the design phase:

4.1.2   Provide a description selecting the <u>sample</u> for the specific statistical activity (irrelevant in the cases where no sample is drawn or for domains based entirely on administrative data). Confirm that the steps defined in sub-process 2.4 were followed. If not, describe any differences from the design phase:

> **4.1.2.1 Is there any special tagging system in place to limit the burden on data providers prior to sampling, panelling, etc.? (E.g. exclusion of households already included in other surveys or the same survey of previous years):**

> **4.1.2.2 If yes, describe the system in place:**

4.1.3   The setting-up of any mailing databases should be described here, together with panelling, over-sampling or rotation:

4.1.4   Mention (and possibly quantify) any quality issues pertaining to the sampling frame, e.g. under/over coverage, duplicate records, time lag, etc.:

> **4.1.4.1 Under-coverage (units that should be included in the frame when in fact they were not, i.e. missing in-scope units):**

> **4.1.4.2 Over-coverage (units included in the frame when in fact they shouldn't be, i.e. out-of-scope units that should not be included in the target population (e.g. due to wrong NACE or size group)):**

4.1.4.3 Duplicates:

4.1.4.4 Any other quality issues (e.g. time-lag between reference period and latest update of frame etc.):

SUB-PROCESS 4.2: 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. 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. This sub-process includes:

- preparing a collection strategy;
- selection and training of collection staff;
- ensuring collection resources are available e.g. laptops;
- agreeing terms with any intermediate collection bodies, e.g. sub-contractors for computer assisted telephone interviewing
- 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 etc.).

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

4.2.1   Describe the organisation of data collection (include briefing sessions):

4.2.2   Describe the process of training the staff - including coders, data-entry persons, enumerators and supervisors. E.g. duration of training, share of technical and content training, separate training for supervisors and interviewers, training material used etc.:

4.2.3   Describe how work is distributed among survey interviewers in workloads (e.g. households assigned to interviewers prior to data collection):

- 31 -

4.2.4   Specify how sampling units are informed about their inclusion in the survey:
☐Letter
☐Email
☐Phone call
☐Other (please specify):

4.2.5   Explain the recruitment process for enumerators and supervisors:

4.2.6   Describe any problems encountered due to the current recruitment process:

4.2.7   Describe the criteria applied in assigning supervisors and enumerators:

4.2.8   Describe the process of ensuring that the necessary resources are available (e.g. netbooks/laptops, paper questionnaires):

4.2.9 Describe the procedures followed for preparing the collection instruments (e.g. printing questionnaires, loading questionnaires and data onto enumerators' netbooks/laptops, pre-filling them with existing data etc.):

4.2.10 Describe any measures taken regarding the security of the data collected. E.g. whether interviewers need to take an oath, encryption and backup of electronic data collected, use of password protected netbooks/laptops, instructions for safety precautions when handling paper-questionnaires (returning catalogues used and completed/draft questionnaires, shredding, etc.):

4.2.11 Describe any agreements with sub-contractors regarding ways to collect data:

4.2.12 For administrative sources, describe the procedures applied in order to ensure that the necessary processes, systems and confidentiality procedures are in place, for receiving or extracting the necessary information:

SUB-PROCESS 4.3: RUN 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. 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. This sub-process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints. For administrative and other non-statistical sources, this process is brief: the provider is either contacted to send the information, or sends it as scheduled. When the collection meets its targets, 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.

4.3.1 For surveys, describe the fieldwork procedures applied including:
- Follow-ups, reminders, any legal letters sent
- Treatment of any undelivered mail (e.g. to find alternative addresses etc.).
- Auditing of interviewers' fieldwork, analysis of unit/item non-response by interviewer
- Handling of any refusals, complaints and queries

4.3.2 For administrative data, describe the procedures followed, including:
- How and when are data providers contacted to provide the data
- Any basic checks made on the structure of the files received/quick validation of data (if any). E.g. are the files received in the right format? Do they contain the expected fields? More laborious checks fall under sub-process 5.3.

SUB-PROCESS 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 take-on, for example using clerical staff or optical character recognition 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.

4.4.1   In the case of paper questionnaires:

4.4.1.1 Describe the data-entry process:

4.4.1.2   Describe archiving/shredding procedures of material taking place once the data entry is finalised:

4.4.2   In the case of electronic questionnaires, describe the process of joining and exporting data files:

4.4.3   Describe any discrepancies between planned (stated in sub process 1.6) and actual data collection costs (e.g. in man-days):

4.4.4   Describe any checks that are carried out after all data has been keyed-in and ready for analysis (e.g. structure of files, import of all variables, format of variables etc.).This does not refer to the auditing of interviewers' work which is included in sub-process 4.3:

PHASE 5 – PROCESS

Process							
5.1 Integrate data	5.2 Classify & code	5.3 Review & validate	5.4 Edit & impute	5.5 Derive new variables & units	5.6 Calculate weights	5.7 Calculate aggregates	5.8 Finalise data files

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. It may be repeated several times if necessary. 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. Activities within the "Process" and "Analyse" phases may 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.

SUB-PROCESS 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 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. There may also be several instances of data integration in any statistical business process. Following integration, depending on data protection requirements, data may be anonymised, that is stripped of identifiers such as name and address, to help to protect confidentiality.

5.1.1   Describe the process of integrating data collected through different <u>collection modes</u> (e.g. data via email, paper questionnaires, electronic questionnaires, pdf forms, web-questionnaires etc.). Include information about the tools used:

5.1.2   Describe the process of integrating data received from <u>administrative sources and/or other surveys</u>. Include information about which administrative sources are used (e.g. Social Insurance, VAT, Registrar of Companies, Civil Registry etc.) and refer to any linkage variables (e.g. VAT number, Social Insurance Number, ID card number, TIC etc.). Include information about the tools used:

5.1.3   In the case of conflicting sources (e.g. different employment data from Social Insurance, LFS, Employment and SBS) explain how the final decision on which data sources to use was taken:

SUB-PROCESS 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 classification scheme.

5.2.1   Describe the procedure followed in order to code the data collected. Include information about the tools used (e.g. software, dictionaries etc.):

5.2.2   The above description should cover all code lists and classification systems provided in sub-process 2.2. If there were any deviations from the original design please describe:

SUB-PROCESS 5.3: REVIEW AND VALIDATE

This sub-process examines data to try 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 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. 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 web collection. Whilst this sub-process is concerned with detection of actual or potential errors, any correction activities that actually change the data are done in sub-process 5.4.

5.3.1 **Describe the various stages of the validation procedure also taking into account the following:**
- The actual imputation and editing phases are dealt with in sub-process 5.4
- Explain how outliers, item non-response and miscoding are examined
- Make reference to complex checks on administrative data (not basic checks included in sub-process 4.3)
- Include information about the tools used

SUB-PROCESS 5.4: EDIT AND 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 (i.e. is erroneous data flagged or is it replaced accordingly in the dataset?);
- 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.

5.4.1 **Describe the procedure for treating incorrect, missing or unreliable data (include information on whether different versions of files are kept). Include information about the tools used:**

5.4.2   If applicable, describe the imputation method used (refer to which variables this is applied and the imputation rate for these variables and whether "old" and "imputed" data are both kept in the dataset.). Include information about the tools used:

5.4.3   Refer to administrative sources that are solely used for editing and to treat item non-response:

5.4.4   Provide information on unit non-response (include reasons for non-response, values etc.):

5.4.5   If applicable, provide information on how outliers identified in the validation phase are treated. Include information about the tools used:

SUB-PROCESS 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.

5.5.1   Describe the method used to derive new variables and units as well as any assumption made. Include information about the tools used. Indicate any deviations from the list of derived variables provided in sub-process 2.2:

SUB-PROCESS 5.6: CALCULATE WEIGHTS

This sub process creates weights for unit data records according to the methodology created in sub-process 2.5 (Design processing and analysis). 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.

5.6.1   Describe the procedure in place to calculate weights providing information on design weights, adjustment for non-response and calibration, where applicable:

5.6.2   If calibration is used, describe the procedure in detail (e.g. calibration variables, tool used for calibration, source of data used for calibration, function used (logit, linear bounded etc.), stratification used etc.):

5.6.3   If upon validation/benchmarking (included in sub-process 6.2) inconsistencies are found and the problem lies within the weights, describe here how these were amended to correct for the inconsistencies.

This sub process creates aggregate data and population totals from micro-data or lower-level aggregates. It includes summing data for records sharing certain characteristics, determining measures of average and dispersion, and applying weights from sub-process 5.6 to derive appropriate totals. In the case of sample surveys, sampling errors may also be calculated in this sub-process, and associated to the relevant aggregates.

- Any aggregate data prepared for benchmarking purposes, namely totals, measures of location (mean, median, etc.), and measures of dispersion (Coefficient of variation, standard deviations, etc.) must be included here.
- Includes also any variance estimation for validation purposes such as: confidence intervals and sampling errors that are calculated at aggregate level for internal checking purposes only (it is important to note that reference must be made to the method not the actual value for the error). It is not very common in practice to calculate these measures at such an early stage but it may happen, e.g. in HBS the confidence interval for income is calculated and compared with that for SILC for consistency.
- It is important not to include here any sampling errors which are calculated and also released – these are to be included in sub-process 6.5.
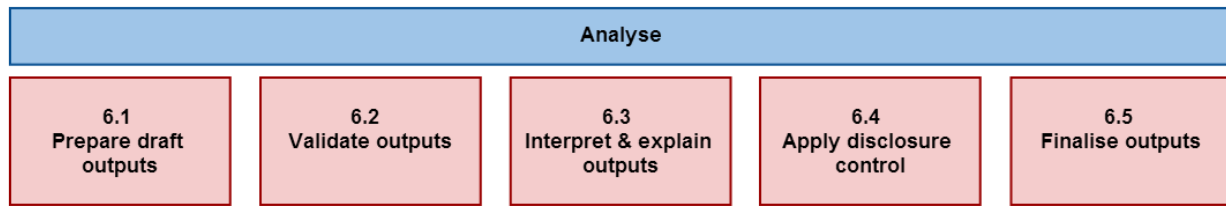
SUB-PROCESS 5.8: FINALISE DATA FILES

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.

5.8.1   This sub-process targets the degree of closeness of computations or estimates to the exact or true value that will ultimately be considered as final in the 'Analyse' phase (e.g. HICP flash estimate). Any significant discrepancies must be described and if acceptable, evidence must be provided to justify them. If these are not acceptable, other sub-processes in the 'Process' phase must be re-addressed until preliminary results are considered satisfactory:

5.8.2 Describe the data files (number of files, structure, format etc.) which are considered ready to be used for data analysis and preparation of outputs:

PHASE 6 – ANALYSE

Analyse				
6.1 Prepare draft outputs	6.2 Validate outputs	6.3 Interpret & explain outputs	6.4 Apply disclosure control	6.5 Finalise outputs

In this phase, statistical outputs are produced, examined in detail and made ready for dissemination. It includes preparing statistical content (including commentary, technical notes, etc.), and ensuring outputs are "fit for purpose" prior to dissemination to customers. This phase also includes the sub-processes and activities that enable statistical analysts to understand the statistics produced. 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.

SUB-PROCESS 6.1: 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.

6.1.1   Describe the procedure to produce the draft set of outputs (e.g. weighted totals, means, median, indices, trends, seasonally adjusted series, working day adjusted series, etc.). This information may include other tables that will not be used in the final publication. Explain the methods/methodology used in calculating/producing the outputs. <u>Include information regarding the tools used</u>:

SUB-PROCESS 6.2: 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. This sub-process also 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 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.

Note: Any revision of weights after the validation of output (e.g. benchmarking etc.) should be included under sub-process 6.3.

Explain how the following activities are carried out in order to validate the quality of the output:

6.2.1   Checking that the population coverage and response rates are as required:

```



```

6.2.2   Checking that the associated metadata are present and in line with expectations:

```



```

6.2.3   Comparing the statistics with previous cycles (if applicable):

```



```

6.2.4   Checking coherence of the statistics with other relevant data (both internal and external):

```



```

6.2.5   Investigating inconsistencies in the statistics:

6.2.6   Performing macro editing:

6.2.7   Validating the statistics against expectations and domain intelligence:

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.

6.3.1   Give details on how the outputs are interpreted and explained. E.g. here one may need to explain the validity of the change in series:

SUB-PROCESS 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. This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques. The degree and method of disclosure control may vary for different types of outputs, for example the approach used for micro-data sets for research purposes will be different to that for published tables or maps.

Note: If outputs need to be revised due to confidentiality issues, these must be specified in sub-processes 6.1, 6.2, etc.

6.4.1   Describe the anonymisation process applied, including information regarding the tools used. Also, provide information regarding residual risks of disclosure (in case risk functions are used), risky combination of variables (in microdata) as well as suppression rates and number of primary and confidential cells:

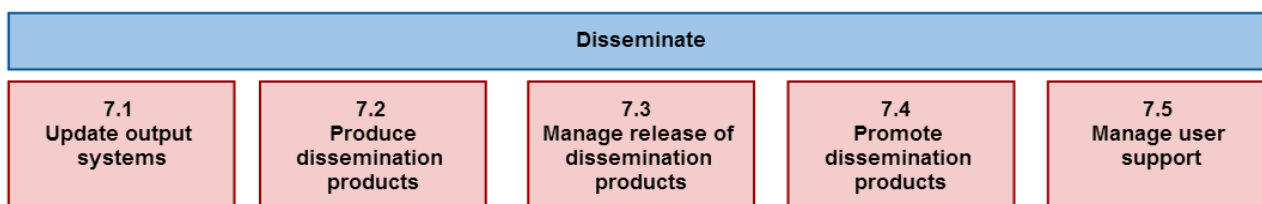## SUB-PROCESS 6.5: FINALISE OUTPUTS

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.

6.5.1   Describe the procedure for finalising the outputs (before the preparation of the dissemination products). This may include the following:
- completing consistency checks;
- determining the level of release, and flagging the data;
- 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.

PHASE 7 – DISSEMINATE PHASE

Disseminate				
7.1 Update output systems	7.2 Produce dissemination products	7.3 Manage release of dissemination products	7.4 Promote dissemination products	7.5 Manage user support

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. It is made up of 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:

SUB-PROCESS 7.1: 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.

SUB-PROCESS 7.2: PRODUCE DISSEMINATION PRODUCTS

This sub-process produces the products, as previously designed (in sub-process 2.1), 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.

SUB-PROCESS 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.

SUB-PROCESS 7.4: PROMOTE DISSEMINATION PRODUCTS

Whilst marketing in general can be considered to be an over-arching 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 web sites, wikis and blogs to facilitate the process of communicating statistical information to users.

SUB-PROCESS 7.5: 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.

PHASE 8 – EVALUATE PHASE

Evaluate		
8.1 Gather evaluation inputs	**8.2** Conduct evaluation	**8.3** Agree an action plan
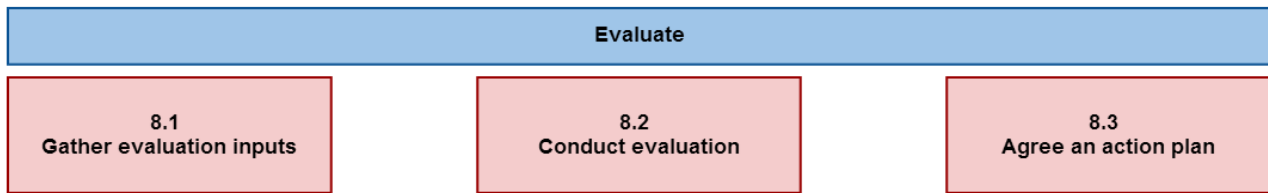
This phase manages the evaluation of a specific instance of a statistical business process, as opposed to the more general over-arching process of statistical quality management described in Section VI. It logically takes place at the end of the instance of the process, but relies on inputs gathered 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.

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 may 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).

This phase is made up of three sub-processes, which are generally sequential, from left to right, but which can overlap to some extent in practice. These sub-processes are:

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

SUB-PROCESS 8.2: CONDUCT EVALUATION

This sub-process analyses the evaluation inputs and synthesises them into an evaluation 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.

SUB-PROCESS 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.