BHAS

Agency for statistics of Bosnia and Herzegovina

Generic Statistical Business Process Model - GSBPM

GUIDELINES FOR ITS IMPLEMENTATION IN THE AGENCY FOR STATISTICS OF BOSNIA AND HERZEGOVINA

Sarajevo, 2018

CONTENT

Sub-process 5.8: Finalisation of data files

PHASE 6 - ANALYSIS

Sub-process 6.1: Preparation of draft outputs

Sub-process 6.2: Analysis of relevancy and validation of outputs

Sub-process 6.3: Interpretation of outputs

Sub-process 6.4: Protection of confidential data

PHASE 7 - DISSEMINATION

Sub-process 7.1: Update of statistical outputs

Sub-process 7.2: Production and presentation of release of statistical products

Sub-process 7.3: Managing release of dissemination products

Sub-process 7.4: Promoting dissemination products

Sub-process 7.5: User support

PHASE 8 - EVALUATION

Sub-process 8.1: Gather documentation about survey

Sub-process 8.2: Conduct evaluation

Sub-process 8.3: Action plan for improvement

I. INTRODUCTION

1. One of the main goals of official statistics is to reduce the costs of statistical production and improve data quality. This requires standardisation of activities, uniformity of the mechanism of production, as well as adjusting work to reduce management errors. That is why the Generic Model of the Statistical Business Process (GSBPM) can be considered as a flexible tool for the production of official statistics and identification and explanation of the process.

2. The GSBPM describes and defines the set of business processes needed to produce official statistics. It provides a standard framework and harmonized terminology to help statistical organisations to modernise their statistical production processes, as well as to share methods and components. It can also be used for integrating data and metadata standards, as well as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement.

3. Following the example of many other national statistical offices which have successfully implemented the GSBPM and have based its quality management system on it, the Agency for Statistics of Bosnia and Herzegovina has also decided to use GSBPM to record the processes of statistical production monitoring. Documentation of statistical processes in the Agency will bring many benefits, such as:

- clarity and transparency of monitored statistical production processes;
- achieving standardised and harmonised procedures through analysis and improvement of existing procedures;
- drafting guidelines for the quality of statistical processes and
- improving overall efficiency of statistical production.

4. The GSBPM phases define the general model framework. Therefore, it is possible that the main processes and sub-processes are not the same for all outputs. Furthermore, phases that are specific to some outputs may not be relevant to other outputs or activities. Thus, the current model consists of 8 major phases, 36 sub-processes.

5. 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.

6. GSBPM is not a rigid framework in which all steps must be followed in a strict order. Some sub-processes will be re-examined several times by forming iterative loops, especially within the phase 'Process' and 'Analysis'.

7. More information on GSBPM-u can be found on UNECE website at the following link: https://statswiki.unece.org/display/GSBPM/GSBPM+v5.0

PHASE 1 – IDENTIFICATION OF NEEDS

This phase is triggered when needs for new statistics are identified or a review of current statistics, based on obtained feedback, is initiated. It includes all activities associated with engaging users to identify their detailed statistical needs, proposing solution options and preparing business cases to meet these needs.

IDENTIFICATION OF NEEDS		
1.1 Identify data needs	1.2 Check and identify data sources – data availability	1.3 Prepare and submit business case

Sub-process 1.1: Identify data needs

Determining needs for data is initiated when the data still do not exist or when the existing data do not sufficiently meet all users' needs for data, or there are new requirements of lawmakers or signed agreements. When determining data needs, a comprehensive and systematic approach is needed, with the involvement of all interested users as well as taking into consideration requirements of the EU regulations. Through the process of determining data needs, we learn what users expect from the competent statistical institutions and which of their expectations regarding data can be met. For this purpose, the process should comprise consultations with all interested users, through various forms of cooperation. It also includes consideration of practice in other international statistical organisations which produce 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.

Sub-process 1.2: Check and identify data sources – data availability

Before reaching a decision to introduce a new statistical survey or redesign a current one, it is necessary to review all existing surveys (already being carried out), as well as the content of administrative sources, and determine first - whether these statistical data are already available in existing sources. If available, the next step is to examine whether the existing sources are compatible with needs for the new survey i.e. whether there are restrictions that prevent the use of these sources to meet new data requirements (differences in methodology, periodicity, purpose of data collection, etc.). Reviewing and researching data sources are primarily done to make a decision on whether - and to what extent - administrative sources can be used as a direct data source. If, in reviewed and available sources, there are no statistical data required by users, it is necessary to redesign existing survey, or to introduce a new survey.

Sub-process 1.3: Prepare and submit business case

To carry out statistical survey successfully, it is needed to plan well the necessary human and material resources, and to set a timeline for performing all tasks. Precise planning of resources and determining a list of activities and deadlines is a key to efficient implementation of statistical activity. Each statistical survey should be included in the Annual Work Plan (programme of statistical surveys). The survey that is currently being set up is included in the Work Plan as a pilot survey, and a survey planned to be implemented regularly, is included in the Work Plan as a regular survey.
Before the start of statistical survey, it is necessary to prepare a staff and financial plan, and timely it in budget preparation as well as to plan the need for additional resources. The costs and benefits that will result in the collection of data should be 'weighted".

Documenting the findings of the other sub-processes in this phase in the form of a business case to begin implementation of a new or modified statistical business process. Such business case usually would comprise the following elements:

- A description of the 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 solution with details 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.

PHASE 2 – DESIGN

This phase describes the development and preparation activities, and any related practical research work needed to define the statistical outputs, concepts, methodologies, instruments for collecting data and operational processes. This phase specifies all relevant metadata, ready for use later in the statistical business process, as well as quality assurance procedures. It is important to use international and national standards in preparatory activities in order to reduce the length and cost of preparatory process, and enhance the comparability and usability of outputs. This phase is broken down into four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative.

DESIGN			
2.1 Design outputs of statistical survey	2.2 Preparation of methodology for collecting data and conducting survey	2.3 Preparation of data sources for creating sampling frame	2.4 Preparation of methodology for statistical data processing

Sub-process 2.1 Design outputs of statistical survey

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. Here are also designed processes that govern access to any confidential outputs. Outputs should be designed to follow existing standards wherever possible, so inputs to this process may include metadata from similar or previous collection, international standards, and information about practices in other statistical offices. This sub-process is aimed at designing the timetable and content of the outputs (statistical tables, indicators, special publications, etc.) planned for dissemination. Outputs design follows internally established procedures and dissemination policy.

Sub-process 2.2 Preparation of methodology for collecting data and conducting survey

This sub-process includes development of all necessary methodologies, (methods, instruments for data collection, variables, definitions, descriptions, instructions, agreements and MoUs with data providers, questionnaire content, dissemination plans, etc.). The preparation of metadata description of collected and derived variables and classifications is a key prerequisite for the next phases. We should use (as far as possible) harmonised concepts, variables and classifications, but the definitions of these must be adapted to specific needs. All definitions of concepts, variables and classifications should be documented as well as any deviations from the recommended standards. The most appropriate methods and data collection instruments are determined. Activities depend on

the methods of data collection (CAPI, PAPI, CATI, and CAWI) including testing instruments. All formal agreements on the delivery of data (e.g. technical protocols) are made.

Sub-process 2.3 Preparation of data sources for creating sampling frame

In this sub-process we specify the target population. Before defining a sampling frame, it is necessary to precisely specify the target population (persons, households, dwellings, farms, enterprises ...) as follows: in terms of the characteristics we are interested in (specify who or what the target population is and what conditions the unit should meet to be part of the target population); the geospatial location of the unit should be determined; a reference period (time) in which we are interested in population characteristics is to be determined. As the main source, we choose the one that allows us to better cover the target population. The main source of observation units is complemented with supplementary sources. The main data source for business statistical survey is, as a rule, a Statistical Business Register (SBR), which contains a list of all enterprises. For surveys where data on persons and households are collected i.e. for survey questionnaires used in field interviews, the main data source is Census of Population, Households and Dwellings. The essential step in implementing the process of preparation of a sampling frame is to establish a procedure for selecting the units to be included in the sampling frame. We then try to "improve" a list of units determined at the specific timeline using additional administrative and statistical data sources.

Sub-process 2.4 Preparation of methodology for statistical data processing

Statistical data processing comprises all procedures taken after completion of the data collecting phase or data taking over, with the aim that the final statistical outputs reflect the characteristics of the observed population as accurately as possible. However, during the planning phase of the survey, first of all, it is necessary to foresee which procedures are important for implementing a planned survey, which methodological procedures should be used and which programme tools would be most suitable for the implementation. This can include specifications of procedures for coding, editing, imputation, estimation, integration, validation and finalisation of data sets with deadlines and responsible persons. This can include design of specifications for IT department, referring to implementation of specific methods for coding, validation rules, imputation coverage etc. Technical details and detailed description should be provided in appropriate sub-processes of the Phase 5, and not here.

PHASE 3 – BUILD

This phase establishes and tests the production solution to the point where it is ready for use in the "live" environment. The outputs of the "Build" phase direct the selection of processes, instruments, information and services configured in this phase to create the complete operational environment to run the process. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, (following a review or a change in methodology or technology), and not for every iteration.
This phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative.

Build				
3.1 Build collection exchange channels and instruments	3.2 Establish software support	3.3 Build dissemination components	3.4 Test data collecting and processing tools	3.5 Test and configuration of statistical business process

Sub-process 3.1: Build collection exchange channels and instruments

This sub-process describes the activities to build the collection exchange channels to be used during the "Collect" phase. A collection may use one or more modes to receive the data (e.g. personal or telephone interviews; paper, electronic or web questionnaires, etc.). Collection exchange channels 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 exchange channel (e.g. testing the questions in a questionnaire). In this sub-process technical specification for building or updating data collection instruments are prepared. Specifications are based on the outputs of the sub-processes of the “Design" phase, where the most appropriate and required instruments are determined. The specifications include a type of instruments that will be used, instructions for defining specific quality control mechanisms such as logical checks, skipping between questions, etc. This Sub-process is conducted on the basis of consultations with the IT department and developers. In cases where administrative or other data sources are being collected, the specifications may be consulted with the owners of the data sources in terms of services and procedures. In this Sub-process, collection instruments have been developed or updated through cooperation between methodologists of the concerned department and IT department or external development. In addition, the testing of the developed or updated software for the collection of statistical data is performed. This is usually done by IT staff and methodologists of the concerned department. In some cases, data entry staff or selected respondents can be involved in testing software for data collection.

Sub-process 3.2: Design of software support

This sub-process configures the workflow and systems used within the statistical business processes, from data collection through to dissemination. It describes actions to be taken for the developing new and improvement of current software components required for business process and designed in the preparation phase.
Components can include control tables and reports, databases, output tables, data transformation tools, data and metadata management tools. This implies the preparation of a technical specification for the establishment or updating components for data processing and data analysing. Specifications are based on the outputs from the Sub-process 2.4 phase "Preparation of methodology for statistical data processing". This sub-process is based on consultations amongst IT department, methodologists responsible for respective statistics and responsible persons for sampling (if relevant). In this sub-process, software for data processing or analysing is developed through collaboration with methodologists, IT (or external programmers), respective statisticians and statisticians dealing with sampling (if relevant).

Sub-process 3.3 Build dissemination components

This sub-process describes the activities to build new components or reuse existing components and services needed for the dissemination of statistical products as designed in sub-process 2.1 "Design outputs of statistical survey". All types of dissemination components and services are included, from those used to produce traditional paper publications to those that provide web services, (linked) open data outputs, geospatial statistics and geographies, maps, or access to micro-data. It is necessary to prepare technical specifications to build or update dissemination components. Specifications include a list of statistical outputs to be disseminated, type and functionality of dissemination tools, rules and standards for visualisation, link to quality and metadata reports and other. This sub-process is conducted on the basis of consultations between methodologists of respective statistics, dissemination staff and IT. It is necessary to develop or update dissemination software.

Sub-process 3.4: Test data collecting and processing tools

This includes testing and sign-off of new programmes and procedures. It also includes testing of interactions amongst components ensuring that production system functions as a coherent set of components. In this sub –process, the initial test of developed or upgraded processing and analysing components are done. It includes collecting data for pilot surveys to test instruments for data collection. Then, processing and analysis of the collected data follows. After pilot surveys, it may be necessary to go back to previous steps and make adjustments.

Sub-process 3.5: Test and configuration of statistical business process

Configuration of the workflow refers to the process from collecting data through archiving statistical outputs. These activities comprise:

- Producing documentation about the process components, including technical documentation and user manuals;
- Introducing and training the business users on how to operate the process. The goal is to introduce the users of production system with its structures, procedures and instructions for the work with its components. The introduction is performed through specific technical and other training as well as the development and use of user manuals and other technical documentation. These activities are often carried out with the support of IT department;
- Moving the process components into the production environment and ensuring they work as expected in that environment.

PHASE 4 - COLLECT

This phase collects all necessary data (data and metadata), using different collection modes (including extractions from statistical, administrative and other non-statistical registers and data bases) and loads them into the appropriate environment for further processing. Although 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 of iterations.
The "Collect" phase is broken down into four sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative.

Data collecting			
4.1 Create frame and select sample	4.2 Set up data collection	4.3 Run data collection	4.4 Data entry

Sub-process 4.1: Create frame and select sample

The sampling frame represents the physical realisation of target population units, the characteristics of which would describe the statistical output. In the next step, it is necessary to select the units (between the units within the sample) that will really be included in the survey, and for which we want to obtain data at later stages. Due to practical (primarily expert) reasons, we rarely include all units from the sampling frame in a survey. It is usually necessary (using an appropriate procedure) to reduce the list of units from the sampling frame to an appropriate size, which still allows us to calculate sufficiently accurate results for the entire population. When the sampling frame is prepared and when a sample size is determined, we begin the procedure for selecting sampling units, which was established at the planning stage of the survey. For this we need the appropriate program environment and special algorithms

Sub-process 4.2: Set up data collection

This sub-process ensures that the people, processes and technology are ready to collect data and metadata. 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.

This sub-process comprises:

- Drawing up a plan for data collection. The plan establishes structures responsible for all activities and organisational units or specific staff (methodologists) who will provide methodological support to survey interviewers or respondents. The plan also defines the deadlines for the fieldwork and reporting by using software for data loading into central database.
- Selecting staff to collect data. The sub-process includes the preparation of individual contracts and communication with staff. In the case of e.g. business statistics where

respondents/reporting units directly submit data to the central database, the responsible contact person is identified. This contact person must check and update the list of respondents' contact lists.

- Preparing training materials. Training materials such as guidelines and methodological notes for interviewers and respondents are prepared. Prepared materials are presented at the training of interviewers or are sent directly to respondents/reporting units in case of collecting data without interviewers. In some cases, training sessions may include the presentation and testing of new or updated data collection methods / input of CATI methods.
- Ensuring material resources and IT instruments for data collections (questionnaires, instructions, methodical notes, official letters and declarations) that are envisaged in sub-process 2.2, and distributing them to relevant structures or actors in the collecting process.
- Ensuring access to data entry programs. In this sub-component, an access to data entry programme is ensured by providing user name and password. It includes installation of software for data collecting/loading data on mobile devices (tablets) or personal computers when necessary.

Sub-process 4.3: Run data collection

Collecting data from reporting units (data respondents) is carried out with the different instruments for data collection. It includes the initial contact with reporting units (data providers) and any subsequent follow-up or reminder actions. The process includes monitoring of timelines and timing of receiving reports in line with the planned schedule. Preliminary database is prepared in this sub -process. Responsible department and statisticians have an enabled access to internal IT tools to examine the preliminary data collected. In addition, the validation of the collected data is also performed. Some basic checks of the structure and integrity of the information gathered may take place within this sub-process, (e.g. checking that files are in the right format and contain the expected fields). In most cases these automated validity procedures are performed by IT department, and results are sent to respective departments for evaluation and further procedures. In case of identified technical problems and following additional measures taken, reporting units are contacted to re-send their reports. This sub-process can include manual data entry at the point of contact or fieldwork management, depending on the source and collection mode. It records when and how data providers were contacted, and whether they have responded. It implies that management of providers is included, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints. For administrative or other non-statistical data sources, this process is short: the provider is either contacted to send the information or sends it as agreed in line with the schedule.

Sub-process 4.4: Data entry

In this sub-process collected and validated data are prepared for processing and analysing. If needed, the data are transferred to relevant databases.

This sub-process includes manual loading of the data from paper questionnaires or with the help of optical character recognition tools, automatic data capture from online questionnaires and applications, or the conversion of data files received from other institutions into another format.

It may include manual or automatic data capture, e.g. using clerical staff for manual data capture or optical character recognition tools to extract information from paper questionnaires or converting of data files formats received from other organisations.

It may also include analysis of the metadata (paradata) associated with collection to ensure the collection activities have met requirements. In cases where there are physical collection instruments, such as a paper questionnaire, which is not needed for further processing, this sub-process manages the archiving of that material. All materials related to the collection activities are archived. This sub-process involves archiving of all materials and instruments used, as well as documenting the performance and activities undertaken to support respondents or interviewers during the collecting. It also includes documentation of the issues encountered during the evaluation of the collected data.

PHASE 5 – PROCESS

This phase describes the cleaning of data and their preparation for analysis. It is made up of sub-processes that integrate, 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 (Weighting), 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 also commence before the "Collect" phase is completed. This enables the compilation of provisional results where timeliness is an important concern for users, and increases the time available for analysis.

Process							
5.1 Integration of various data sources	5.2 Coding	5.3 Review and data validation	5.4 Editing and imputation	5.5 Production of derived variables and units	5.6 Weighting	5.7 Calculation of aggregates	5.8 Finalisation of data files

Sub-process 5.1: Integration of various data sources

This sub-process integrates data from one or more sources. The input data can be from a mixture of external or internal data sources, and a variety of collection exchange channels, 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 data from different sources when such data relates to the same unit, with the aim of linking micro and 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 removed identifiers such as name and address, to help to protect confidentiality.

Sub-process 5.2: Coding

Coding (classifying) includes the assignment of alphanumeric key text response and is usually automated process. Partial manual coding is also necessary for special cases. In order to code the text response automatically, the necessary assumption is that there is a sufficiently developed encryption key. All text responses that cannot be automatically coded should be, if possible, coded by professional. No entry outside of the allowed range of values may be allowed. Through double-entering of the text responses and the subsequent comparison of the entries, the incorrect entries can be reduced to a minimum. The automatic coding parameter should be selected to indicate on the one hand as many text responses as possible, but on the other hand to minimize the number of incorrect coding. The first, it is necessary to determine the ranges of values for each character. In addition, it has to be decided which modalities or which values are used for cases with no numerical characters (e.g. missing values “non-response”, etc.).

Sub-process 5.3: Review and data validation

Review and validation includes logic-calculation data control according to the set control rules, procedure for control of aggregated data or group of units, control of extreme values, outliers, and critical values. Control, editing and correction of data can be carried out more than once until the data do not reach satisfactory level of quality. During the process of formulation of the rules for acceptability at the micro level, the following should be recognized: values beyond the defined range of values; inconsistent and unacceptable combinations of values and missing values.

Sub-process 5.4: Editing and Imputation

For macro editing as many comparing references should be included. Particularly comparisons in relation to previous periods. Process of comparison and editing includes automatic data editing or activating the alert that data should be checked and edited manually. Missing or unreliable data could be replaced with estimated data. It includes the following: identification of possible errors; selection of data that will be included or excluded from imputation; imputation by one or more predefined methods; record of imputed data into the dataset and flagging them as imputed; production of metadata on imputation process. Calculated weights could be used to increase the results from survey based on sample to be representative for the target population or for the adjustment of the non-response rate in the case of inclusion of the entire target population into the sample. The result of the imputation process should be analysed, evaluate and determine – if the desired effects were achieved by the imputation.

Sub-process 5.5: Production of derived variables and units

This sub-process derives variables and statistical units (using arithmetic formula to one or more existing variables that are already present in the dataset) that are not explicitly required in the process of data collection, but necessary for delivery of required results.
New statistical units may be derived by aggregating or splitting data for collection units or by various other estimation methods (e.g. deriving households where the collection units are persons, or enterprises where the collection units are legal units). New, derived variable or units are necessary for calculation of aggregates or calculation of indicators, indices or other statistical results.

Sub-process 5.6: Weighting

Weighting is carried out according to the methodology created in sub-process 2.4 “Preparation of methodology for statistical data processing”. Calculated weights can be used to "gross-up" data from the sample survey to make it representative of the target population or for adjustment of non-response rate in case of including the entire target population in the sample, then for the implementation of new variables or units (sub-process 5.5). Sampling experts are included in this sub-process. In addition, calculated and created weights in this sub-process are integrated into data processing base for further calculations or calculation of aggregates. Sampling experts are included in this sub-process in many cases.

Sub-process 5.7: Calculation of aggregates

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. Calculation of aggregated data can be performed by requesting primary database using a special tool that calculates aggregates or calculates new aggregates derived from the initial ones.
All data collected for comparison purposes should be included here (totals, mean values, median, coefficient of variation, standard deviation, etc.). It also includes estimation of variation for validation purposes, such as: confidence intervals and errors in sampling calculated at the level of aggregates but only for internal checking (it is important to specify the method rather than actual value of the error). In practice, it is not usual to calculate these measures at such an early phase, but, for example, the following can happen: in HBS the confidence interval is being calculated for income and is compared with the same one from the SILC, for consistency.

Sub-process 5.8: Finalization of data files

This sub-process brings together the results of the other sub-processes in this phase in a data file (usually 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. The data files should contain micro data, weights and aggregates and data processed in the sub-process 5.5.

PHASE 6 - ANALYSIS

In this phase, statistical outputs are produced, examined in detail, and prepared for dissemination. It includes preparing statistical content (including comments, 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 data and the statistics produced. The "Analyse" phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced. The "Analyse" phase is broken down into five sub-processes, which are generally sequential, from left to right, but can also occur in parallel, and can be iterative.

Analysis			
6.1 Preparation of draft outputs	6.2 Analysis of relevancy and validation of outputs	6.3 Interpretation of outputs	6.4 Protection of confidential data

Sub-process 6.1: Preparation of draft outputs

This is sub-process where the data are transformed into statistical outputs. It includes production of additional measures such as indexes, trends or seasonally adjusted series, as well as the recording of quality characteristics. Quality indicators such as non-response rate, sampling errors or other necessary indicators are also calculated in parallel with the preparation of draft outputs. Quality indicators could be used for further checks of the results in the sub-process 6.2, or as additional information in sub-process 8.2. In addition, in this sub-process, the calculation of seasonally adjusted time series is performed.

Sub-process 6.2: Analysis of relevancy and validation outputs

Analysis of relevancy of outputs is a process where the “meaningfulness” of outputs, the internal consistency, coherence and comparability in time and space, and compliance with the existing internal and external referent data sources is checked. The analysis of relevance, i.e. the confirmation of the outputs, is performed at the macro-level in the process of data editing and includes the following procedures: checking the internal consistency of the results, for example: Checking outputs based on known or expected links between results (if it is for example value of production higher than the value added); Verification of compliance of outputs comparing the results from previous reference periods (firstly, this is valid for those surveys for which the primary purpose is not to measure changes in time); Checking the compliance of outputs with related or linked outputs from other statistical surveys (outputs from the surveys conducted by statistical institutions in B&H, or outputs from the surveys conducted by other institutions); Internal verification of outputs in statistical institutions and occasional verification of "relevance and meaning" of outputs with external experts.

Sub-process 6.3: Interpretation of outputs

In this sub-process we must consider the method of data collection (coverage and source: sample-based survey, administrative database) and other information related to obtaining data (e.g. non-response rate). All of this, together with appropriate metadata, should be provided to the user. In particular, all data insufficiencies should be described, such as e.g. the deviations between the target population and the population that we observed. Principle of Data Confidentiality should be also considered in the data interpretation. When interpreting the outputs in the index form and other relative numbers, a reasonable period of comparison should be selected, which will allow the user of the data to effectively interpret the movement of any occurrences. Interpretation of the outputs must be adjusted to the target population and the media (users) in which the data will be later published. Interpretation of results must be impartial, objective, accurate and understandable. During the presentation of outputs in the form of index and other relative numbers, one should be careful in order to correctly interpret changes when occurrences are expressed in percentage (%) or in percentage points. NPR. If in 2015 in comparison with 2014, the number of households that have PCs increased from 21.3% to 23.3% - this increase does not mean 2%, but 2 percentage points or 9.4%.

Sub-process 6.4: 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 disclosure, as well as the application of the use of techniques to prevent or obstruct the access to data or suppression or perturbation techniques (perturbation is a method of data modification that involves a slight change of data in order to reduce the detection risk, while retaining as much content and structure as possible). 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 (microdata).

PHASE 7 - DISSEMINATION

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

Dissemination				
7.1 Update of statistical outputs	7.2 Production and presentation of release of statistical products	7.3 Managing release of dissemination products	7.4 Promoting dissemination products	7.5 User support

Sub-process 7.1: Update of statistical outputs

Publication of statistical data is performed according to a standardized procedure and in different technologies. Standardized procedures are based on predefined structures, formats and metadata, and they are considered in the process of preparing the table (in the statistical data processing phase). For each form of publication of statistical data and information, the output data should be updated in advance; these can be data in the form of a database or in the form of tables with final aggregated data that are ready for publication. This sub-process manages the system updates in which the data and metadata are stored for the purpose of dissemination, including the following: formatting data and metadata that are ready to be included in the database with outputs; loading of data and metadata into the database with outputs and linking data with relevant metadata. Formatting, loading and linking of metadata should take place in earlier phases, but this sub-process includes final check that all of the necessary metadata are ready for dissemination.

Sub-process 7.2: Production and presentation of release of statistical products

This sub-process produces the products that are previously designed (2.1 Design outputs of statistical survey) to meet user needs. It can include printed publications, press releases and web-sites. Elements of the content of the publication should be prepared considering the target users and the purpose itself. The general principles for presenting the results differ according to the type and publication media. The user's statistical literacy and different understanding of statistical data and information should be taken into account for data

presentation. The data should be presented in a way that is understandable and interesting to the user.
The data, which are published in the first electronic publication - the release and the popular publications, should be presented as a statistical story that should contain the comment and visualization of the data. The comment must be effective - short, simple, understandable and interesting. Data visualizations must be created and summarized in a clear form with simple tables and graphs. In principle, the text should explain everything that is necessary for understanding the presented tabular outputs. This also includes a brief description of the concepts and methods used. Detailed methodological explanations and metadata are given in the Quality Reports for individual statistical surveys.

Sub-process 7.3: Managing 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. Correction of errors is the activity that is also related to the management of publishing statistics. The purpose of correction of errors in the published statistical data is to provide accurate and quality statistical data and information for users. The error correction system must be designed to be clear and understandable to users.

Sub-process 7.4: Promoting dissemination products

Whilst marketing in general can be considered as 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, as well as the use of tools including web sites, and blogs to facilitate the process of communicating statistical information to users. The best way to get feedback from the user is to test their satisfaction. Based on feedback, it is possible to perform an analysis and evaluation of user needs.

Sub-process 7.5: User support

User support includes the provision of statistical data and information, counselling on access to data, and on the use of data-processing tools, and assistance in the search and preparation of statistical data and information. The user is also enabled the access to statistically protected microdata, but under special conditions and a special contract. User support modes may vary depending on which type of request is being made (written request for data, telephone inquiries, access to statistically protected microdata, and use of the library).

PHASE 8 - EVALUATION

Evaluation		
8.1 Gather documentation about survey	8.2 Conduct evaluation	8.3 Action plan for improvement

Statistical surveys are usually carried out periodically, so that the entire statistical process is repeated. It is important that this process involve a feedback link, which enables the introduction of changes and improvements. For these purposes, each statistical survey (after it is completed) needs to be fully evaluated, the success of the entire survey should be critically assessed and possibilities for improvement identified. Collecting information on the quality of statistical data takes place during the entire statistical process. Systematic documenting of individual phases of the survey is an important part of the information on the process of the survey and helps in identifying potential systematic errors in this process. With this information we can evaluate the quality of statistical data and critically evaluate the results obtained that are important for users, as they gain additional insight into the data collection process. Publication of information on data quality is a transparent way of informing users.

Sub-process 8.1: Gather documentation about survey

The process of preparing survey documentation includes a detailed description of the statistical activity, including a description of the concepts, definitions, methods used, the used information system, and work instructions. The quality of survey documentation is an important indicator of the quality of survey. It is also an important tool for communicating between different surveys, as well as between producers and users of statistics. The survey documentation is part of the metadata. Documentation can generally be divided into documentation for users of survey results and documentation for producers of statistical survey. *The documentation for the users of the survey results* describes and documents the statistical results and is publicly announced. The purpose of the documentation for the users is to present data and the statistical methods used for measurement, increase the understanding of the data itself, make it easier for users to request data and, if necessary, further process them. Examples of documentation for users are: survey questionnaires, methodological guidelines, user-oriented reports on the quality of statistical research. *The documentation for the produce of statistical surveys* describes the statistical procedures and processes that are used in the entire statistical process. This documentation is created during the implementation of certain steps in the statistical survey process and is mainly intended for internal use. The documentation for producers of statistical surveys must, first of all, describe in detail the steps undertaken in carrying out the survey (e.g. targeting the population, sample selection, designing questionnaires, editing data, publishing data, etc.).

Sub-process 8.2: Conduct evaluation - evaluation inputs

This sub-process analyses the evaluation inputs, compares them to the expected/target benchmarking results (when available), and synthesises them into a quality report. The quality report should take into account all quality issues that are specific to the statistical business process, should highlight the deviation of the measured performance from the expected values, and should make recommendations for changes if necessary.

The evaluation can take place at the end of the whole process (ex-post evaluation) for selected activities, allowing for quick fixes or continuous improvement.

Sub-process 8.3: Action plan for improvement

This sub-process brings together the necessary decision-making power to form and agree an action plan based on the evaluation report. It includes steps or actions for follow up of recommendations trough proposed (and established) mechanism at institutional level.

QUALITY MANAGEMENT/METADATA MANAGEMENT							
1 Identification of needs	**2 Design**	**3 Build**	**4 Collect**	**5 Process**	**6 Analysis**	**7 Dissemination**	**8 Evaluation**
1.1 Identify data needs	2.1 Design outputs of statistical survey	3.1 Build collection exchange channels and instruments	4.1 Create frame and select sample	5.1 Integratation of various data sources	6.1 Preparion of draft outputs	7.1 Update of statistical outputs	8.1 Gather documentation about survey
1.2 Check and identify data sources – data availability	2.2 Preparation of methodology for collecting data and conducting survey	3.2 Establish software support	4.2 Set up data collestion	5.2 Coding	6.2 Analysis of relevancy and validation of outputs	7.2 Production and presentation of release of statistical products	8.2 Conduct evaluation
1.3 Prepare and submit business case	2.3 Preparation of data sources for creating sample frame	3.3 Build dissemination components	4.3 Run data collection	5.3 Review and data validation	6.3 Interpretation of outputs	7.3 Managing realease of dissemination products	8.3 Action plan for improvement
	2.4 Preparation of methodology for statistical data procesing	3.4 Test data collecting and processing tools	4.4 Data entry	5.4 Editing and imputation	6.4 Protection of confidential data	7.4 Promoting dissemination products	
		3.5 Test and configuration of statistical business process		5.5 Production of derived variables and units		7.5 User support	
				5.6 Weighting			
				5.7 Calculation of aggregates			
				5.8 Finalisation of data files			