REPUBLIC OF CROATIA

CROATIAN BUREAU OF STATISTICS

MAIN (STATISTICAL) BUSINESS PROCESSES

INSTRUCTIONS FOR FILLING OUT THE TEMPLATE

CONTENTS

INTRODUCTION

The template is a result of analysis carried out within the Croatian statistical system, in a way that some of individual phases of the statistical business are adjusted to the needs like collection, processing, publication and dissemination of statistical data.

The mentioned template will be used for describing, documenting every statistical survey and determining the level of quality. The template was created according to the Generic Statistical Business Process Model - GSBPM and is intended for all activities undertaken by producers of official statistics at national and international level. GSBPM is adjusted to the needs of the Croatian statistical system, designed as a model independent of the data source so that it can be used for the description and evaluation of the quality of the processes based upon surveys, censuses, administrative records and other non-statistical or combined data sources.

The statistical business process usually involves the collection and processing of "raw" data for producing statistical results. GSBPM is applicable to cases where existing data are revised or time series re-calculated. In such cases, the input is taken as a previously published statistical data which are then processed and analysed for obtaining the revised results (outputs). In such cases there is the possibility of skipping several sub-processes and some work stages.

Beside that it is applicable to processes that provide statistical results, GSBPM can also be applied to the development and maintenance of statistical registers in which the input is similar to the statistical production (although in such cases a greater emphasis is put on the administrative data), and the results are usually frames of data set which are then used as input for other processes.

GSBPM is not designed as a strictly defined framework in which all steps must be applied in the same order, but as a model that identifies steps in a statistical business processes and interdependencies between them. Although the presentation of the model shows the logical sequence of phases of work in most statistical business process model elements can appear in a different order. Therefore the GSBPM is a simple model and therefore widely applicable.

The model includes the following activities:

– The Quality Management - This process includes mechanisms for evaluation and quality control. Recognizes the importance of evaluation and feedback during the statistical business processes.
– The Metadata management - Metadata are created and handled within each phase and, therefore, there is a strong requirement of the metadata management system to ensure that the appropriate metadata retain their links with information during the application GSBPM.
– Statistical framework management - It includes the development of standards, such as methodologies, concepts and classifications applicable through many processes.
– Statistical programme management - It includes the systematic monitoring and control resulting information requirements for these liabilities and changing data sources across all statistical areas. This can result in the definition of new statistical business process or redesign of existing ones.
– Knowledge management - It ensures repeatability of statistical processes, mainly by maintaining processing documentation.
– Data management - This includes process-independent aspects of the process, such as general data security, custody and ownership of data.
– Process data management - It includes management of data and metadata derived from all parts of the statistical business processes and providing information about them.

– Provider management - This includes managing the burdens that are transferred from process to process, as well as topics such as profiling and data management for the contact (and thus closely related to the statistical business processes that maintain registers)
– Customer management - It includes general marketing activities, promoting statistical literacy and treatment of non-specific customer feedback.

1. Specify Needs

1.1 Determine needs for information and necessary results

The sub process includes further development of the existing statistics as well as the initial investigation and identification of necessary new statistics and what is needed of statistics, monitoring of national legislation, relevant regulations of the European Union and the research of the customer needs. It is necessary to have a good knowledge about the customer needs and ensure mutual understanding in order to know what the statistical organisation must not only deliver, but also when, how and, perhaps most importantly, why. During the second iteration and all subsequent iterations of this phase the main task will be to determine whether the previously identified needs are changed.

1.2 Consult & confirm needs

It is checked whether the current internal and external data sources meet the demands of users and the conditions in which they are available. Evaluation of possible alternatives includes research of potential administrative data sources and associated methodologies to determine whether they can be used for statistical purposes. Includes assessment of the legal framework within which data will be collected and used, and preparing suggestions for changes of existing legislation or the introduction in a new legal framework.

1.3 Establish output objective, analysis and testing possiblities

The statistical results are determined, needed to meet user needs identified in the sub process. The information needs and the required results are defined. The targetresults and concepts for the introduction of new and / or revision of existing resources with the implementation of feasibility study are defined. The analysis results are documented in order to obtain approval for the use of new or adjusted statistical business processes.

Results of analysis include:

– A description of the existing business process (if one exists) information on how the current statistics are produced, with a focus on inefficiencies and issues to be addressed
– The proposal for future solutions that explains in detail how the statistical business process will be developed that will produce a new or revised statistics
– Assessment of costs and benefits as well as any external constraints.

2. Preparation and development of statistical methodologies

2.1 Definition and development of the methodology for collecting data and conducting survey

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

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

2.2 Defining a framework and methodology for the sample selection

Identifies and defines the target population, defines sample frame (and, where necessary, the register from which the sample is taken), and determines the most appropriate criteria and sampling methodology (which may include the entire target population). The most common sources are geospatial, administrative and statistical registers, as well as censuses. This sub process describes the ways in which these sources can be combined. The analysis should be carried out to check whether the sample frame covers the target population. The sampling plan should be developed: a sample is made in sub process of the sample selection by using the methodology defined in this sub process.

2.3 Development of methodology for data processing

The development of methodology for data processing involves determining procedures for coding, editing, imputing, estimating, integrating, linking geospatial and statistical information, validating and finalizing the data sets with the deadlines. The general rule is to check first the existing solutions in order to determine whether the purpose of this process and re-use through a variety of statistical business processes is appropriate. This sub process also describes who will be responsible for what and when.

3. Build necessary instruments for enforcement

3.1 Build data collection instrument

Preparation of project requirements includes requirements for data collection and processing with the deadlines in accordance with the instructions for making the IT project requirements.

3.2 Build instruments for data collection

It includes the creation of instruments that will be used in the phase of data collection. The instrument for data collection is created or produced on the basis of design specifications made in the preparation phase. It is recommended that the connection of instruments with a system of statistical metadata and the ability to save paradata[1)] after data collection.

Possible instruments for collecting statistics are:

– on-line form (CAWI)
– hard copy form
– media for download data (data files, internet services, protocols, etc.)
– WEB scraping
– CAPI
– CATI.

3.3 Configure workflows

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

3.4 Testing instruments for data collection and data processing

It includes technical testing and approval of new programs and procedures. Includes testing the interaction between the components and ensures that the production system is functioning as a coordinated set of components. Includes data collection for experimental studies in order to test instruments for data collection. This is followed by processing and analyzing the collected data. After experimental studies it might be required to return to previous steps and make adjustments.

1) Paradata are process metadata that are collected during data collection and processing (e.g., the method of data collection, the duration of the interview, etc.).

3.5 Test statistical business process

Configuration of the production process flow refers to the data collection until archiving final statistical results. It includes activities that process puts into production ready to be used in business areas. These activities include:

- production of documentation about the process components, including technical documentation and user manuals
- training of business users on how to manage the process
- moving the process components in a production environment and ensuring that these components are functioning as expected in this environment

4. Data collection

4.1 Selection of final population/sample

The selection of the target population for the survey including design and preparation of the required address lists. For surveys that are conducted on sample the frame is determined and the sample is selected for data collection, as specified in sub process Defining a framework and methodology for the sample selection. Includes sample coordination within the statistical survey (for example, management of overlap or rotation) and between different statistical surveys that use the same frame or register (for example, management of overlap or equal distribution of the response burden). Quality assurance, maintenance of frame and selected samples are also performed in this sub process. Sampling in this sub process is not usually important for surveys based on the use of pre-existing data sources (e.g. administrative data). After sample selection follows the preparation of draft address lists and all other ancillary activities.

4.2 Preparation of data collection

This ensures that the staff, processes and technology are ready for data collection, in all the modes as designed. This sub process includes:

- preparing a plan for data collection
- training of collection staff
- providing resources for data collection, such as laptops
- ensuring the data confidentiality for data that will be collected
- preparation of the instruments for data collection (e.g. addressing the statistical questionnaires, announcement letter, questionnaire printing, collation, pre-filling of questionnaires with existing data, loading questionnaires and data onto interviewers´ computers, etc.)
- creating an application for data collection (CAWI, CAPI, CATI), web scraping.

4.3 Primary data collection

It conducts data collection using various instruments to collect data. It includes initial contact with data providers and any activities related to subsequent follow-up or reminder actions. This sub process captures when and how the data provider was contacted and whether the data provider responded. When the data collection fulfils its goals (which usually manifests in response rate) the data collection is concluded and report on the collection is drawn up, such as response rates.

4.4 Overtaking data from administrative and other secondary sources

Download of the administrative data are based on contracts made between institutions, while the data transfer takes place via different media (optical media, network protocols, register replications, etc.). During the data download the control of structures as well as data formats are performed.

4.5 Entering of data collection

It includes entering of collected data and metadata for further processing in the process. It may involve manual data entry from printed questionnaires or using an optical reader, automatic downloading of data via Internet questionnaires and web scraping applications (CAWI, CAPI, CATI) or converting data files received from other institutions in other format.

5. Data processing

5.1 Integration of data collection

It includes integration of data from one or more sources for data processing. The input data (input) can come from a combination of external and internal data sources as well as from different ways of data collection. The result is a harmonised data set. The integration of data includes:

- procedures for matching / record linkage to connect data from different sources when such data refer to the same unit (e.g. geospatial data with statistical or other data)
- prioritising, when two or more sources contain data for the same variable (with potentially different values).

Data integration can be done at any time of data processing. After integration, depending on the needs of data protection, data can be anonymised, i.e. the identifiers can be removed such as name and address in order to protect confidentiality. It includes the classification and coding of the input data.

5.2 Control, editing and data correction

Control includes input data validation by defined control rules, process control of aggregated data or unit groups, control of extreme values, outliers, critical values. The process of editing and correcting includes automatic data editing or activation of warning that should manually review and correct the data. Control, editing and data correcting can be performed repeatedly until the data do not reach a satisfactory level of quality.

5.3 Imputation and weightening

The missing or unreliable data can be replaced by the estimated data. These include:

- identification of potential errors and gaps
- selection of data to be included or excluded from the imputation
- imputation using one or more of predefined methods
- entering the imputed data back to the data collection set and flagging these data as imputed
- production of metadata about the process of imputation.

Weighting is carried out according to the methodology drawn up in subprocess Defining a methodology for statistical data processing. Calculated weights can be used to gross-up the results of the survey sample, that would be representative on the target population, or to adjust the rate of non-response in the case involving the entire target population in the sample.

5.4 Production of derived variables

Production of derived variables and statistical units that are not explicitly collected during data collection using the arithmetic formulas in one or more existing variables. New statistical units can be derived by aggregating or dividing data collection for units included in collection or by different assessment methods (e.g., performing household, when the units included in data collection are persons or companies, when the units are legal persons).

5.5 Calculating the aggregate

From the micro data the aggregated data and population totals are calculated (e.g. aggregation of data according to economic, social, geographical and other classifications). Data are summarized data by the common characteristics, the measures of average and dispersion are determined and weights from sub-process Imputation and weighting are applied. If the method of random sampling is applied, the standard errors can be calculated.

5.6 Calculation of final data files

The results of all sub-processes are combined at this stage, resulting with final data file used as input in the analysis phase. Sometimes it is an intermediate, not a final data file, especially for statistical surveys that are very important and where there is a need for the production of preliminary and final estimates.

5.7 Production and updating registers and database

It refers to data preparation and updating of all registers, nomenclatures and classifications, and the production of output database.

6. Analyse

6.1 Statistical analysis of results

The collected data are transformed into statistical results. It includes production of additional measures such as indices, trends or time-adjusted series as well as the recording of quality characteristics. In order to increase the value and create prerequisites for analyses of statistical data, this sub-process could include the preparation of maps, GIS results and geostatistical services.

6.2 Quality control results

The quality of the produced results is validated, in accordance with the general framework of quality and with expectations. The information that can be applied to a specific area are collected. Quality control includes:

– checking whether the coverage and response rates are in accordance with the expected values
– comparison of statistics with previous cycles (where applicable)
– comparison with other relevant data (external and internal)
– finding inconsistencies in the statistics
– macro editing
– generating Quality Reports for all statistical surveys prescribed by the Annual Implementation Plan

6.3 Detailed analysis and interpretation of data publishing

A detailed analysis and interpretation of results is carried out. Assessed is, how well statistics reflect the initial expectations observing statistics using different tools and media. It ensures that statistics reaches the required level of quality and readiness for use. These include:

– implementation of consistency checks
– determining the level of release and applying caveats
– collecting additional information, including interpretations, summaries, measures of the reliability of results and other necessary metadata
– production of auxiliary internal documents (reports to the Director)
– discussion prior to the publication of results with internal experts for these questions
– approving of the statistical results for publication.

6.4 Protection of confidential data

The protection and confidentiality of disseminated data and metadata are ensured. This may include checking the primary and secondary confidentiality as well as application techniques for preventing or impeding access to data.

7. Dissemination

7.1 Design and production of dissemination products

The systems in which data and metadata are stored for dissemination are updated, including:

– formatting data and metadata ready for inclusion in the output database –
– loading data and metadata in the output database
– linking data with relevant metadata.

Publications, press releases and content on the website are produced. The most common steps include:

– preparing the product components (explanatory text, tables, charts, cartographic representations, etc.)
– assembling products form components
– editing products and taking care about respected publishing standards.

7.2 Management of published disseminated products

It includes briefings to specific groups such as press or state administration bodies, as well as dealing with possible restrictions before the release and delivery of products to subscribers.

7.3 Promote dissemination products

Refers to the active promotion of statistical products produced in a specific statistical survey to allow these products to reach the widest group of users. It includes the use of customer relations management tools and the use of tools that include websites, GEOSTAT portal, quality reports and blogs in order to familiarize users of statistical information.

7.4 Manage user support

The customer queries are recorded and responses are provided within the specified time. Such queries should be checked regularly to ensure the input of a comprehensive quality management process because queries may indicate new or changed customer needs.

8. Evaluate

8.1 Gather evaluation inputs

Material for evaluation may result from any phase or sub process. The evaluation may be different, including user feedback, metadata about the process, quality indicators and employee suggestions. The reports on progress with the action plan agreed in the previous period can also be input for the evaluation in the following periods. All these inputs are provided to a person or team that performs quality assessment.

8.2 Conduct evaluation

The inputs for evaluating and synthesising the quality report are analysed. The report should note any problems with the quality of this statistical survey period and, if necessary, recommend changes. These recommendations may be related to changes in any phase or sub-process of future survey periods or may suggest that a certain sub processes are not repeated.

8.3 Agree action plan

In this sub process gathered are responsible person for the purpose of creating and agreeing an action plan based on the quality report. The plan should include evaluation of the production process with suggestions for its improvement.