HUNGARIAN CENTRAL STATISTICAL OFFICE

HUNGARIAN GENERIC STATISTICAL BUSINESS PROCESS MODEL (HGSBPM)

v2.3

Budapest, 2018

Contents

Hungarian Generic Statistical Business Process Model

The goal of the HGSBPM

The main goal of the Hungarian Generic Statistical Business Process Model (HGSBPM) is to make the Hungarian Central Statistical Office's (HCSO) work, its processes more transparent by describing, presenting and interpreting the HCSO data production processes in a harmonised way. The HGSBPM is the model for statistical business processes, by which we, at the HCSO mean the statistical data production process as well as processes related to maintenance of statistical registers and survey frame production. The HGSBPM is also suited for describing other statistical processes like maintenance of classifications, producing questionnaires, data files, publications etc.

The HGSBPM is based on the international Generic Statistical Business Process Model (GSBPM v. 5.0[1]), which is a process matrix referring to the whole statistical data production designed within the UNECE[2] framework. Data production processes do not include functional activities like running the office, economic, legal, HR or management areas or processes. They do include, however, general tasks like quality management, statistical data management (including data protection) and metadata management, which represent the so-called comprehensive areas of data production and extend to the whole of the data production process.

The structure of the HGSBPM

The HGSBPM model incorporates process phases of the statistical data production, sub-processes of breakdown of the process phases', as well as comprehensive areas (quality management, statistical data management and metadata management).

In the HGSBPM – for the sake of practical feasibility – key elements of sub-processes have been highlighted when **process phases** were defined. The HGSBPM is focusing on the process of statistical data production, not on the existing IT applications supporting these, which may serve more sub-processes of more process phases.

The HGSBPM structure of individual sub-process description is:

- Goal of the given sub-process;
- Definition of the main activities in the given sub-process, including sub-process input;
- Actions following the conclusion of a given sub-process and description of the outputs produced.

The content summary description of the **comprehensive areas** in the HGSBPM contains the general characteristic account of every process (ex. goal, main content elements). In addition, the key actions of comprehensive areas appear in the description of individual process phases as well.

[1] See: Generic Statistical Business Process Model GSBPM (Version 5.0, December 2013).
[2] UNECE: United Nations Economic Commission for Europe.

Applying the HGSBPM

Like the GSBPM, the Hungarian Generic Statistical Business Process Model (HGSBPM) is a matrix structure model meaning the sub-processes appearing under the process phases do not necessarily follow each other in a linear way during process fulfilment. Some sub-processes do not appear during certain data production phases, or they can repeat during a certain process, and the represented order of sub-processes does not necessarily show the execution order or direction of the process. The HGSBPM is then a nonlinear model, and by applying it not every sub-process has to be employed. The weight of given sub-processes may differ in the described process as well. The real routine has to be documented with it, which may consist of fewer sub-processes in certain cases than in the case of the comprehensive model. One main goal of the HGSBPM is to allow the description of all statistical business processes of the Office by combining the existing elements of the model.

Practical areas of use of the HGSBPM:

- Methodological standardization: the Office's process review by process phases or sub-processes, developing standardized methodologies based on existing practices, standardization of existing solutions and creating methodological resolutions for potentially missing elements;
- Process documentation: detailed process documentation for specialists along the lines of the HGSBPM, part of it may consist of flowchart production, or certain sub-process activities' documentation based on flowcharts;
- Process monitoring: monitoring of processes, measurement of target values linked to processes, as well as quality measurement, process improvement proposals, possible intervention in the process;
- Process management: systematic process monitoring on system level, fulfilment of assigned targets, target development for missing areas, even starting projects;
- Process optimization: permanent process improvement (even reorganization) in favour of efficiency and success.

Advantages of the HGSBPM:

- By identification of process phases and sub-processes, responsibility areas are defined, interactions between sub-areas, sub-activities can be clearly defined;
- Individual data production processes are described based on a unified concept, which increases the transparency and comparability of the Office's processes. Serves as a base documentation in case of employees' turnover as well;
- Certain capacities are easily aggregated at the process-planning stage;
- Comprehensive areas are default and are included – based on common principles – in certain statistical business process phases; process description makes their later management possible, meaning that by precisely describing the process flow quality check can be performed for any phase, or in case of process modification the new sub-processes can be easily integrated into the existing description.

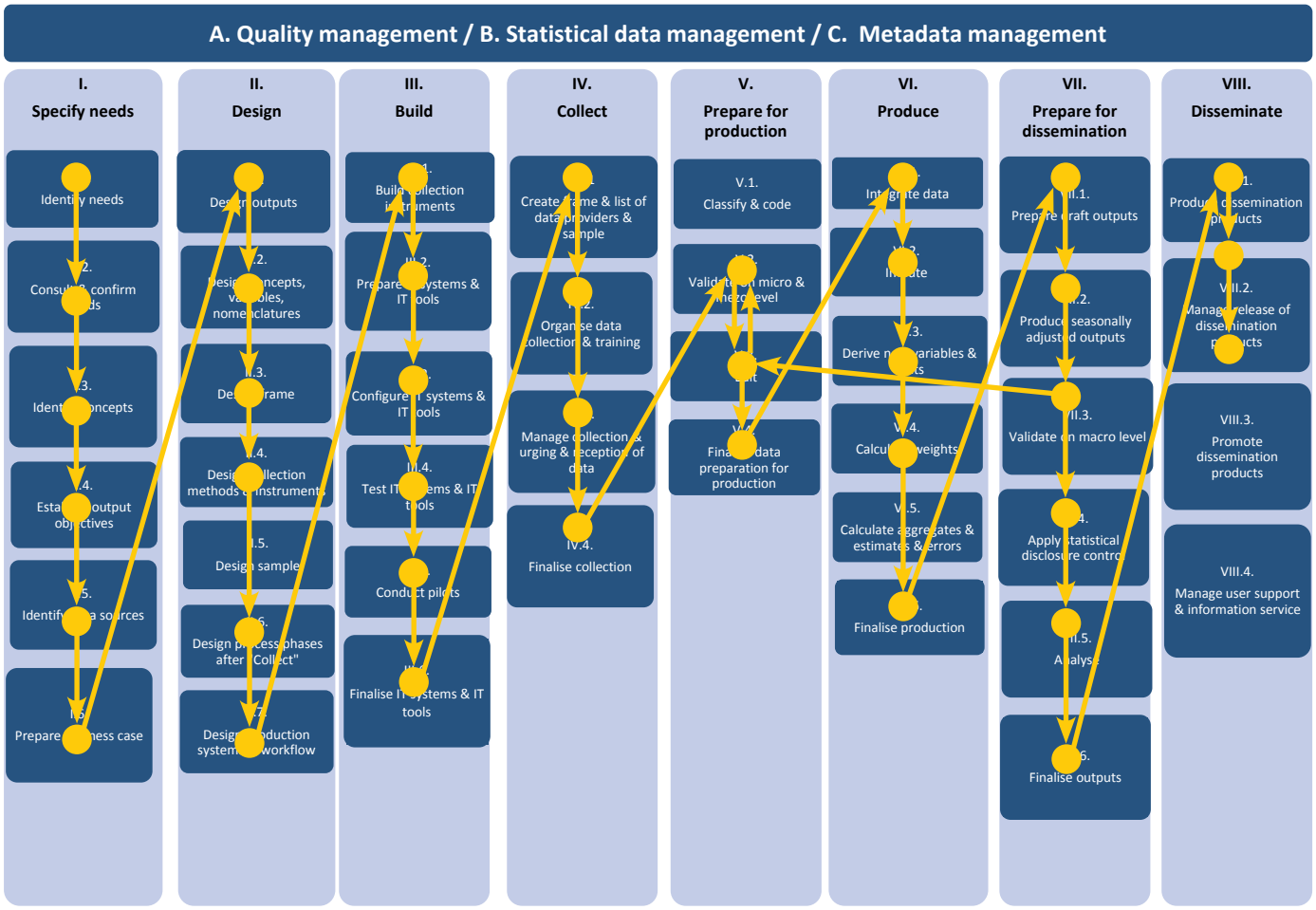


Figure 1: example on using the HGSBPM on a full-scope statistical survey

Reference topics, units of the HGSBPM

When documenting statistical data production processes one has to refer to process phases, sub-processes described in the HGSBPM. Such documentation could be produced

- About planning, fulfilling, primarily processing and publishing certain statistical surveys (data collections and data transmissions)
- In similar concept-structure but on a higher aggregated – e. g. statistical domain – level;
- About planning, maintaining registers and records, about survey frame creation or primary information dissemination based on registers;
- About planning, running general processes some survey descriptions could refer to;
- About secondary regular and occasional data processing and dissemination;
- About other data production processes, e. g. classification maintenance.

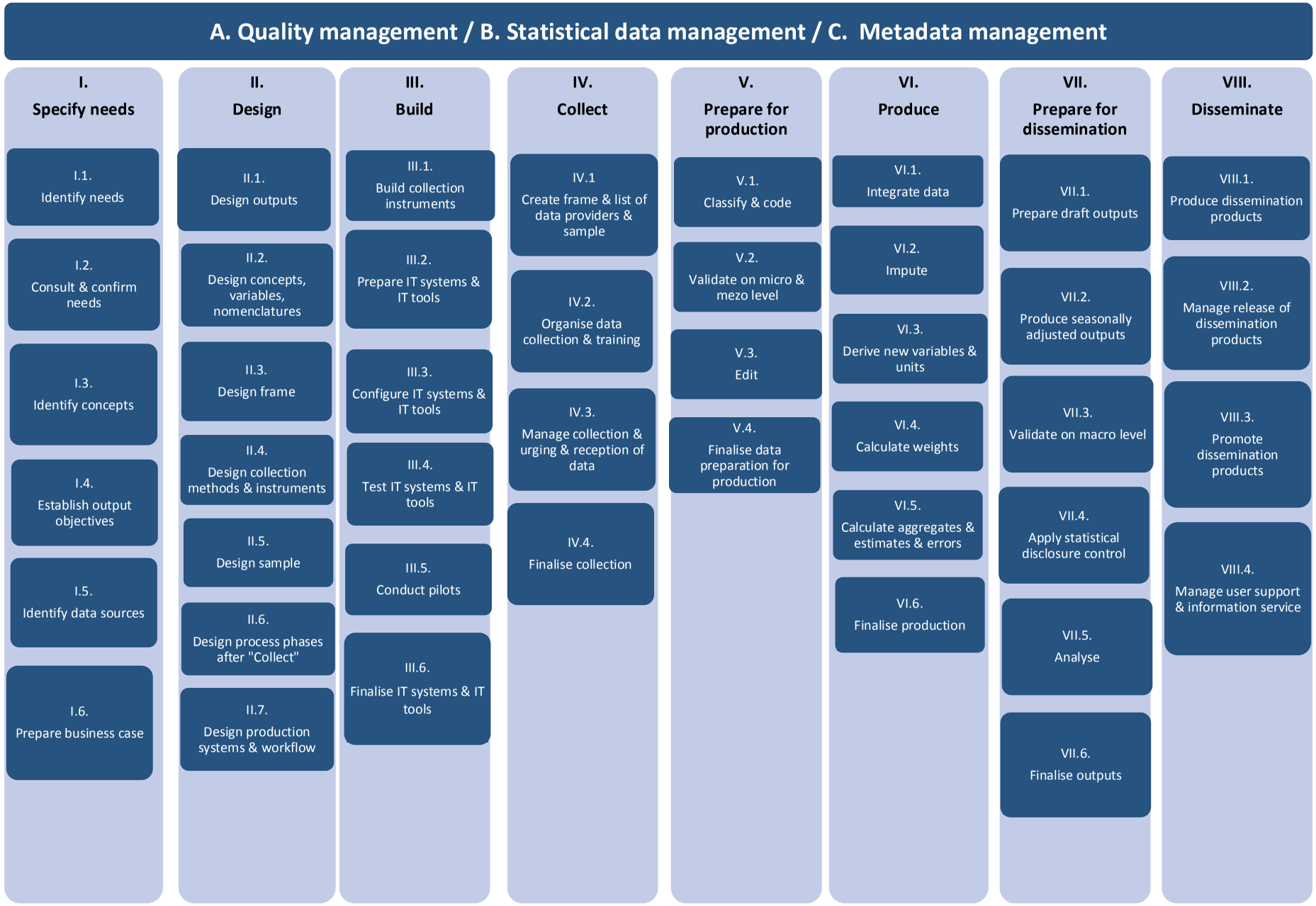


Figure 2: The Hungarian Generic Statistical Business Process Model

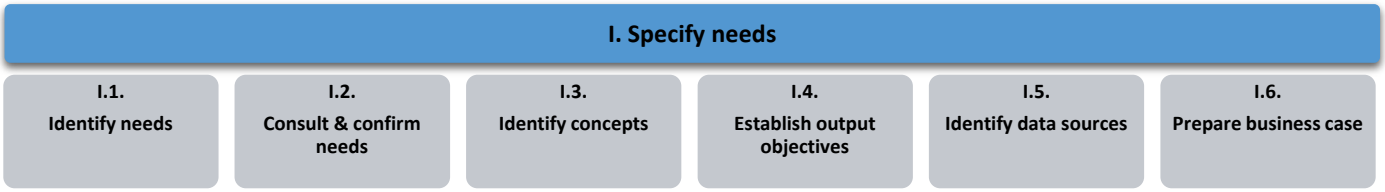
Content description of the Hungarian Statistical Business Process Model (HGSBPM)

The following description contains the explanation of the eight process phases of the HGSBPM (correlating the HCSO process phases with GSBPM v. 5.0's certain process phases) adjusted to the Hungarian specificities.

Process phases and descriptions of corresponding sub-processes are followed by the presentation of the HGSBPM's three comprehensive areas: quality management, statistical data management, metadata management.

Process phases and sub-processes of the statistical data production process:

I. Specify needs



The primary goal of the process phase is to gather and systematize the external (user, data provider) and internal (for realisation of statistical data production processes) needs appearing in the statistical information domain, and to decide about their professional relevance as well as regarding initiation of the detailed planning of data production process based on agreed-upon needs. It can be considered a theoretical preparation phase during which the detailed content items of the occurring external and internal needs are clarified. Contains the revision of present outputs, and if need be, the definition of new outputs, so these could better serve new and changing demands.

We consider a need – beside the expectations regarding production of new statistical information – suggestions regarding modification of existing statistical information, producing more detailed information or quality improvement (e. g. timeliness or punctuality of estimates), changes in information based referencing. Along with an unchanged content of statistical information needs regarding data production process modification, efficiency improvement could show up in this process phase. In the Specify needs process phase the data protection side of data management already appears – it incorporates the confidential treatment of statistical data as well as personal data protection.

In this process phase the statistical organization:

- Makes a survey of – and gathers – needs;
- Defines the target linked to the emerged needs and the corresponding characteristics (concepts, indices);

- Examines if these needs can be served by the primary and secondary data sources of the HCSO or new data sources should be explored;
- Defines, revises the connection between new needs and other statistical subject-matter domains (e. g. reference scope, frequency, applied method's perspective)
- Examines if the statistical information to be produced meets the European Statistical System's requirements and recommendations;
- Makes proposals regarding methods of meeting needs (data collection and data transmission, survey methods, data production processes);
- In case of needs regarding data production process modification examines the possibility of generally applying the needs on other statistical information.

I.1. Identify needs

The scope of the sub-process is to gather, describe new external and internal needs or ones to be modified (irrespectively of whether the person responsible for the data production process is already familiar with the need based on previous experience, or these needs arise during a separate coordination) in order to determine which statistics would be necessary and what needs arise regarding to their quality. Needs could arise on decision maker level, e. g. government and European Statistical System level data needs, it could be based on national or international legal norms, or could arise from other, professional organizations (e. g. societies, chambers, researcher communities) direct user feedback (e. g. user satisfaction survey or user needs survey) or from other statistical subject-matter domains within the Office. Needs revealed during a permanently evaluated quality management by the HCSO in regard to applied data production processes, national or international practices and methodologies belong here. The need means the definition of the factual statistical goal as well; especially in the case of data management of natural persons' personal data. The sub-process closes with the definition of the raised needs.

I.2. Consult & confirm needs

The scope of the sub-process is to harmonise needs with all concerned parties and to make professional decisions based on relevance of needs. In this sub-process user needs are reconciled, ranked, priority lists are created and irrelevant needs are excluded. Reconciliation of needs contains recognition of needs not fulfilled by the present practice, collision of diverging needs of those concerned, the filtering of unnecessary information. Definition of target population (reference scope), reference period and requested statistical data frequency takes place. The precise knowledge of needs helps choose the best method for statistical data production.

Reconciliation of modifications requested in the data production process with the persons responsible for connected process phases, and the examination of data-independent solution methods applicable in data production processes of other statistical information also belongs to the sub-process. The sub-process closes with the definition and prioritizing harmonised needs.

I.3. Identify concepts

The scope of the sub-process is the definition of target population-characteristic criteria (concepts, indices) – which came up in the user needs defined in sub-process I.2 – from users' perspective.

It is not a target on this level to harmonise the identified concepts with the ones used in the present practice (concepts defined based on user needs and those used in present practice are harmonised, and concepts, indices are chosen in sub-process II.2). The sub-process closes with denomination and definition of concepts based on harmonised needs.

I.4. Establish output objectives

The scope of the sub-process is to define the outputs which have to be produced or modified in order to serve user needs defined in sub-process I.2. Quality requirements for outputs have to be defined here considering statistical disclosure control measures applicable to certain outputs. The sub-process closes with the definition of harmonised needs-based outputs and their quality requirements.

I.5. Identify data sources

The scope of the sub-process is the quest of data sources in order to produce information based on adequate concepts, population in the expected quality. All data sources have to be located during the sub-process which would be available for producing statistics according to approved needs (e. g. use of existing statistical information, involvement in statistical data collection, use of secondary data sources). In case the expected data production is possible by new or modifiable data collection, the possibility of serving data collection by multiple sources should be considered along with statistical data domain definition and estimation of necessary capacities. In the sub-process all alternatives for utilising different data sources should be defined – by which the serving of expected needs is ensured.

I.6. Prepare business case

The scope of the sub-process is the preparation of recommendations based on information needs referring to starting new or modifiable surveys or secondary use of available data sources. The survey recommendation proposal facilitates the making of a positive or negative decision about starting the detailed planning of the survey. It contains the emerged information needs, coordinated and approved during needs specification sub-processes, the main concepts necessary for survey completion, outputs, related quality requirements and accessible data sources (statistical data domains) and the necessary resource need for utilizing all these. Acceptance of the survey recommendation proposal makes the start of a detailed plan possible. The sub-process closes in the case of new needs for statistical information with the completion of the survey recommendation proposal, in case of needs for modifying statistical data production processes with preparing requirement specifications.

II. Design

In this phase the statistical data production process is designed; its overall theoretical preparation took place in the previous phase (I. Specify needs) based on user needs. The task in case of a new survey is the build-up of the whole data production process, in case of survey modification the adjustment of the previous, existing data production process. It incorporates:

- Detailed analysis of data sources;
- Defining concepts, nomenclatures utilised in the data production process;
- Defining survey frame;
- Preparing sample design in case of representative statistical surveys;
- Designing methodologies for data preparation, processing, for dissemination preparation and dissemination;
- Designing survey instruments;
- Preparing necessary metadata for survey design;
- Defining outputs for data production processes;
- Designing process work and expense needs.

At the end of the Design stage documentation for detailed survey plan is available, completion can start based on it.

II.1. Design outputs

The scope of the sub-process is the detailed output design defined in sub-process I.4 (e. g. statistical tables, plan for data to be stored, microdata, publication design, homogenous statistical data domains etc.), and examination of the necessary instruments for producing these, including formulation of necessary development needs. In case of personal data output design incorporates data storage in agreement with data protection principles. The sub-process closes with concrete, detailed output design (e. g. logical database design, tables design etc.).

II.2. Design concepts, variables, nomenclatures

The scope of the sub-process is the definition of the content and preparation method of statistical measures in compliance with and to be calculated and published based on concepts formulated in sub-process I.3; the definition of basic statistical measures and variables which will be collected by the Office through different survey instruments. Furthermore the classifications, nomenclatures to be used during data production processes must be defined. In this sub-process the concepts, measures, nomenclatures based on needs will be defined and definitions,

value sets previously attributed to them will be harmonised, corresponded (ex. creation of correspondence tables, measure connections definition). The sub-process closes with defining concepts, measures, nomenclatures necessary for the data production process.

II.3. Design frame

The scope of the sub-process is the definition of target population (reference scope) in conformity with the sub-process I.2 harmonisations; definition of the survey frame covering it (it depends on sampling-based survey the sampling frame) as well as defining the frame's reference period.

The source definition for survey frame configuration takes place in this sub-process (e. g. administrative data sources, statistical registers, records, previous full-scope censuses and representative surveys, maps etc.). The sub-process closes by defining the survey frame.

II.4. Design collection methods & instruments

The scope of the sub-process is the detailed analysis of data sources identified in sub-process I.5 and decision-making in accordance with accepted goals and resource-frames about the survey method definition (statistical data collection, data transmission). It is verified in this sub-process if a corresponding legal framework is available for data source utilisation. In case the legal framework is not available, or not in a satisfactory quality, then an adequate legal basis formation is to be initiated (e. g. co-operation agreement signed or extended with administrative data owner) within the sub-process. In case of voluntary data collection when handling personal data the sub-process contains the design of dissemination and the agreement of those concerned. Design of measuring instruments based on survey method are also part of the sub-process. In case of data collection the questionnaire's content and visual design takes place in this sub-process and the requirements regarding data carriers (laptop, electronic self-loading applications) are gathered here. In case of data transmission: statistical data domain, data structure, form, frequency, receiving deadline, data providers / record keeping units are defined by considering finality and data protection principles. If supporting materials are needed (e. g. filling-in guides, code stocks, reply forms, demonstration instruments, educational resources, data supply motivating instruments, guides, data protection information material etc.), or preliminary guidance, communication is needed for data collection success, then planning of all these is also a part of the sub-process. The sub-process defines the final data sources, datasets, survey methods (statistical data collection, data transmission) of the data collection in question; in case of statistical data collection the collection methods (e. g. paper-based form by post; inquiry by phone, web-based data supply; direct observation etc.); defines the questions in the form, the question blocks, outlines data collection tools. In case of secondary data source transmission agrees with the data owner about the scope of data- and metadata sets', their form, frequency, the reception method (usually laid out in a co-operation agreement). The sub-process closes with the update of the data source registers and with the completion of the data collection and transmission plan.

II.5. Design sample

The goal of the sub-process is the definition of the sampling method and frequency; in case of stratified sampling the definition of stratification criteria, degree of accuracy for estimations to be produced, and the definition of sample size, all based on the survey frame defined in sub-process II.3. It includes consistency provisions between the samples containing identical unit types utilised by different statistical areas as well as the avoiding of overlaps in order to minimise data providers' burdens. The defined sample must be selected in sub-process IV.1. based on methodology set forth in this sub-process. The sub-process closes with the completion of the sampling plan containing the sampling method.

II.6. Design process phases after 'Collect'

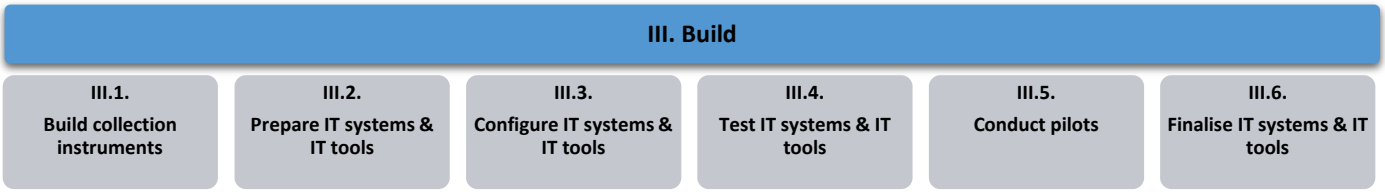
The purpose of the sub-process is provision for data preparation, processing, dissemination as well as planning the methods to be used in dissemination and archiving, which will be utilised in process phases V – VIII and in the comprehensive B area[3]. We include methodological and related IT solutions' planning, specifications. Characteristically, the sub-process contains the definition of data entry methods, encoding, validation, editing, imputation, estimation, error calculation, data integration, seasonal adjustment, statistical disclosure control, in case of personal data anonymisation, pseudonymisation, and planning of every other database finalisation process. The sub-process closes with the completion of the specifications of the concerned sub-processes.

II.7. Design production systems & workflow

The scope of the sub-process is the data-route description, completion of the quality assurance plan, definition of the whole data production process, responsibility order and expense items starting from data collection up to archiving. During survey process compilation the connections between sub-processes, repetition conditions, decision-making points have to be examined. The sub-process contains the persons' access rights to the systems, names of persons responsible, partly responsible, contributors, as well as expense and time needs of these activities, deadlines, risk management. The sub-process contains the preparation of the quality assurance plan, including work-standards, official liabilities and the system of quality control actions for elimination, monitoring and evaluation of sampling and non-sampling errors, as well as for managing different points of the data production process; national level data management of significant quantities of personal data or, in case of applying new technologies data protection impact assessment. The sub-process closes with the compilation of a detailed survey plan, in which the detailed specifications outlined in the sub-processes of process phase II play a part.

[3] See details in „Comprehensive areas of statistical data production".

III. Build



The task of the Build process phase is to prepare planned collection instruments, configure and test technical, IT instruments and systems which form the background of data production processes, to check their compatibility and to put them into operation. It includes:

- Complete data recording of data production processes' metadata;
- In case of data collection the final design and execution;
- Preparation of supporting materials for survey, information and educational materials;
- Carrying out data production processes in IT tools and systems, their testing from data collection to dissemination;
- Testing the whole data production process;
- Finalising the data production process.

This process phase includes the pilot survey extending to the whole data collection process and field work – pilot surveying as well as preparation based on the experiences of tests, of the regional implementation. System documentation and metadata recording is completed in this process phase, based on the detailed survey plan.

At the end of the Build process phase the survey system is ready for implementation.

III.1. Build collection instruments

The goal of the sub-process is the selection, acquisition, development, testing of instruments suitable for the requirements enlisted in sub-process II.4. as well as content testing and finalising of survey measures and supporting materials. Testing of usability of survey methods and their correction – for diminishing data providers' burdens, increasing reliability and the quality of the data to be collected – takes place here. Building safety guarantees necessary for dealing with confidential – including personal – data is also part of the sub-process. The sub-process closes with content finalisation of survey measures and supporting materials.

III.2. Prepare IT systems & IT tools

The goal of the sub-process is the preparation of IT applications' system design based on specifications of the design phase, finalising and recording of corresponding metadata. The sub-process closes with preparation of the system documentation necessary for development, with metadata recording and launching the necessary IT acquisitions.

III.3. Configure IT systems & IT tools

The goal of the sub-process is the completion of the tasks set forth in process phase II and clarified in sub-processes III. 1. and III. 2. through IT instruments supporting the data production process. Databases are created here, the implementation of sub-processes' tasks in IT applications, or new instruments are developed. In case of data collection the actual design of the questionnaires, the implementation of the e-questionnaires with IT instruments are tasks of the sub-process. For the data production process to function, interface-connection between different instruments must be ensured. The sub-process also incorporates the design of production instruments for dissemination products up to all publication forms of paper-based and digital outputs defined in sub-process II. 1., including microdata access. At the end of the sub-process the IT instruments to be used in the data production process are ready to be tested.

III.4. Test IT systems & IT tools

The scope of the sub-process is to perform the necessary tests – according to the task's character – prior to the IT instruments to be used in the data production process are put into operation; e. g. the user functional test (compliance with professional specifications), IT security test, load test, integration test. This sub-process is logically connected to sub-processes III. 2. and III. 3. The sub-process closes with the production of test reports (these reports must include risk assessments, vulnerability enquiries etc.)

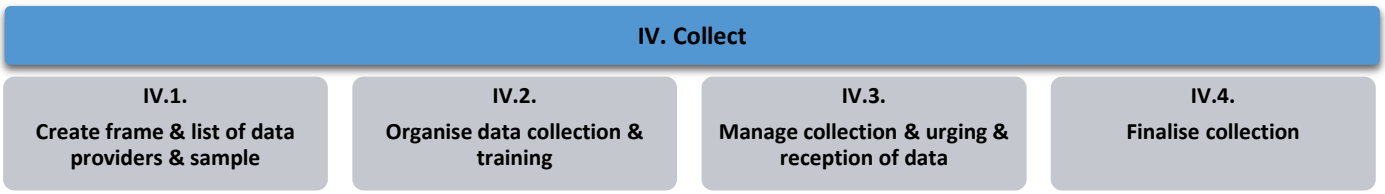
III.5. Conduct pilots

The scope of the sub-process is the testing of the whole data production process. It usually means a survey of a smaller scale than the target population or sample – where the survey methods, its technical support, data processing, analysing etc. can be tested. The sub-process closes with the analysis and evaluation of the pilot survey results, conclusions and their documentation.

III.6. Finalise IT systems & IT tools

The goal of the sub-process is to finalise the IT instruments, to prepare the necessary documentation for their operation. Program execution documents of the IT applications as well as enacting clauses of data collection and preparation are produced and distributed within the sub-process. IT instruments are transferred from the test environment into the operational one. At the end of the sub-process the data production process is ready for use.

IV. Collect



This process phase includes data collection by reference periods according to the completed plans, which could be accomplished, beside direct collection, by combined methods as well (ex. data selection from secondary registers, or statistical registers), includes electronic data access and loading them into the appropriate environment (e. g. into databases). In the sub-process collected data can only be modified by the person who supplied the data. Any other data modification, correction takes place in the Data preparation process phase. It includes:

- Creation of collection frame;
- In case of representative surveys the actual selection of the sample, the definition of data suppliers' scope based on the survey plan;
- Survey organisation;
- Survey-specific skills training;
- Actual data collection;
- Reminders for data suppliers;
- Reception of collected data;
- Recording of data provided by data providers with no corrections;
- Collected data inclusion into databases.

IV.1. Create frame & list of data providers & sample

The scope of the sub-process is, in case of data collection, is to establish and analyse the survey frame based on the statistical registers' (frozen) status at an appointed time and based on other resources (administrative data sources, registers, earlier surveys, maps etc.); in case of data providers, representative surveys the definition of the sample. The sample selection, following the survey frame analysis, will take place according to the design elements assigned in sub-processes II.3. and II.5. In case of data transmission tasks for the given timeframes will be established, broken down on data owner. The completion of the register's status for an appointed time (frozen status) and that of the complete code is not a survey based on registers, it is rather part of the given register's, code system's data production process. The sub-process closes in the case of statistical data collection with the selection of the survey frame, in case of a representative survey with the actual sample selection, in case of data transmissions with the definition of data receiving tasks by data owner.

IV.2. Organise data collection & training

The goal of the sub-process is to ensure human and IT resources set forth in the survey design, as well as to inform data providers and data owners about their data supply duties and invite them to provide data. Accordingly the questionnaire finalised in sub-process III.2. and its supporting materials, directories are produced, mailed; organisers, interviewers participating in the survey are trained. The survey's preliminary communication is also part of the sub-process in case of census-like, primarily household-based data collections. The sub-process closes with the data providers' orientation, and in case of census type data collections with the readiness of the interviewers.

IV.3. Manage collection & urging & reception of data

The goal of the sub-process is to gather statistical – and connected meta- and paradata, to remind data providers, data owners, as well as data reception. Data are provided by individuals or economic organisations, they are collected by methods and instruments established during design, like census (interview, questioning), data supply (paper or web-based declaration), direct observation or data transmission. In case of censuses the process consists of field work and data reception. Accordingly, the coding process – executed by the interviewers – is also part of the sub-process. In case of self-loading surveys we consider contact with data suppliers, data reception and reminders as being part of the sub-process; during electronic data collection data recording takes place at the same time as data collection. At electronic data collections the questionnaire's basic connections are built in the data collection instrument, the data provider has to correct errors prior to sending in the questionnaire. Resending erroneous questionnaires, data files (for correction by the data provider) are part of the sub-process. In the framework of the sub-process takes place the coding and registration of reasons for missing questionnaires or data receptions, their assignment for replacement and sanctions for those refusing to provide data. The sub-process ends by closing data reception.

IV.4. Finalise collection

The scope of the sub-process is the electronic storage of the collected data in the state they were collected. Non-electronically collected data are recorded in their collected state – with no corrections – and stored together with the survey's other collected data in a database. This is how data collection is closed. The sub-process ends with the availability of the data file ready for processing.

V. Prepare

This process phase contains the coding of text fields in the microdata-set ready for preparation, the record-data level verification, correction in order to prepare microdata for processing. Verification of non-electronically collected, recorded data also takes place in this process phase.The data file ready to be processed is the result of the process phase.

The process phase includes:

- Coding of text fields;
- Elementary data verification in order to comply with designed validity rules;
- Labelling elementary-data non-complying with validity rules elementary data (e. g. outliers);
- Correction of erroneous data and marking corrected data;
- Loading ready-for-processing data into standardized databases.

V.1. Classify & code

The goal of the sub-process is to code collected data in accordance with different nomenclatures, classifications, as designed in sub-process II. 6.

This may take place automatically or be performed manually by specialists with the help of IT tools & systems. Within this the text responses from questionnaires or the answers to open-end questions within received data must be corresponded with elements of a given nomenclature, or the coding system in question has to be adopted to the coding system to be used by the HCSO. In case of paper-based surveys the coding activity done by the interviewer during data collection is part of sub-process IV. 3. The sub-process closes with the availability of the encoded micro-data file.

V.2. Validate on micro & meso level

The goal of the sub-process is to study if the available record-level data correspond to the assigned validation criteria. We can use several data sources during micro-validation for examining validity of record-level data. The study always takes place on record data level. Within it coding is verified, record-level data correlation inside or between data sources is examined. Data comparison, verification with data originating from similar or different (e. g. statistical, administrative) data sources, registers or data from previous periods is performed. In this sub-process takes place, among others, the identification of outliers or the handling of erroneous coding. Meso-validation process used in the Office belongs here, which means a more comprehensive correlation–verification based on several surveys than micro-validation is. The sub-process closes with the gathering of discovered errors, their ranking by severity and the availability of the validated microdata file.

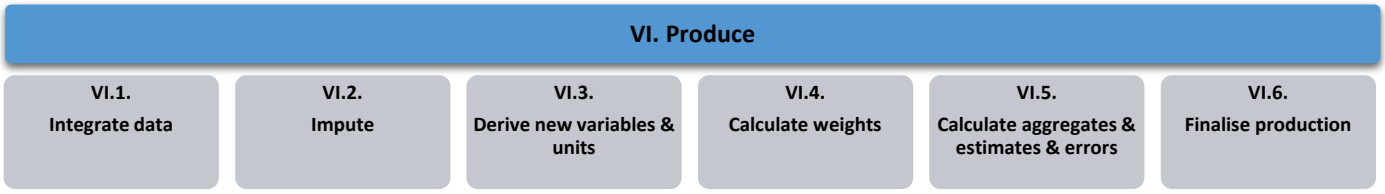
V.3. Edit

The scope of the sub-process is the actual handling of the errors discovered during validation, their correction by defined rules, by automatic or manual methods. Error correction could take place based on corrected questionnaires coming from data providers, enumerators, oral consultation, or in case of their absence based on professional criteria. The sub-process closes with the correction of the erroneous values, labelling corrected data and the availability of the edited microdata file.

V.4. Finalise data preparation for production

The goal of the sub-process is the storage of verified, corrected data along with their labelling. The storage – complying with the database design structure – of data described in the file's metadatabase must be ensured within the sub-process. Data processing is closed in this way. The sub-process closes with the availability of the verified, accepted, ready-for-processing (encoded, micro-validated, edited microdata) data file.

VI. Produce



The scope of the process phase is to produce all data – organised into databases, processed – necessary for preparing dissemination materials as well as serving as inputs for secondary processing. The process phase's inputs are the prepared, data files ready for processing and the results of other processing phases. The process phase includes:

- Matching and linking data files;
- Imputation on unit or item level;
- Creating new measures, indices, new statistical units based on available information;
- Creating weights for estimations;
- Calculating estimations and errors attributed to them;
- Creating aggregates;
- Data labelling (edited, imputed etc.).

The process phase closes with producing the data files containing processed data ready for secondary processing and dissemination preparation.

VI.1. Integrate data

The scope of the sub-process is to link data files coming from one or more sources. The sub-process' input may be an intra- or extra organisational data file or their group or a combination. In case the units of the data sources to be integrated partially overlap and possess unique, correct identifiers or appropriate combinations for key variables, then data integration takes place through these. If we have no knowledge about overlaps between data source units or there are variables,

indices not observed in both data sources then we connect the records referring to similar units by applying common variables. Integration takes data protection principles into account. In case of personal data anonymization belongs here. The sub-process closes by connecting the concerned files.

VI.2. Impute

The scope of the sub-process is to fill the empty cells with data based on assigned cells to be filled in previous sub-processes or according to registrations in sub-process II.6. Replacement (imputation) of unit-level and item-level missing values belongs here. The sub-process closes with the actual replacement of missing values and with labelling replaced data.

VI.3. Derive new variables & units

The goal of the sub-process is to produce variables and statistical units which are not directly available when data is collected (ex. specific indicators, calculated indicators, indices to be produced as process results, index series) but are absolutely necessary for producing predetermined outputs. During the process the necessary indices are produced from the variables in the data file based on predetermined algorithms, processes, rules, according to sub-process II.6. Production of new statistical units could take place by aggregating accounting units or by estimation. An example could be the creation of household-level data when the observed unit of the collection is the individual; or the observed unit is the legal entity, but the information unit is the group of companies. The sub-process closes with the production of new statistical units and measures.

VI.4. Calculate weights

The scope of the sub-process is to produce weights attributed to statistical units. Forming final weights includes considering non-responses, outlier treatment, calibrating, as well as dealing with the variation of weights among years, their conversion (homogenising). Weights are used for producing estimations referring to the whole population. The sub-process closes with the availability of the final weights.

VI.5. Calculate aggregates & estimates & errors

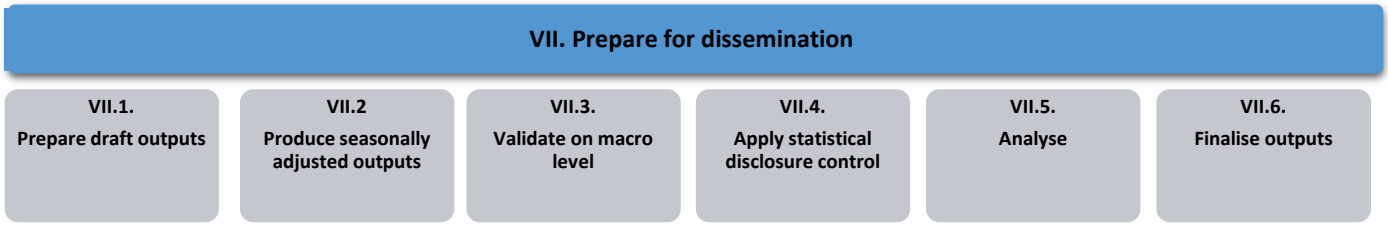
The goal of the sub-process is to model aggregated data and 'total'-s based on record-level (micro) data, produce estimations and to calculate their errors based on methods from sub-process II.6. Aggregation could take place based on some grouping criteria (e. g. by territorial dimension) or in time (e. g. totalling monthly data for quarters, year). Estimations are produced with the help of final weights.

We also calculate estimation errors here. The sub-process closes with the availability of aggregated values, indices, estimations and their errors.

VI.6. Finalise production

The scope of the sub-process is processed data storage (including produces weights, errors and labels). These represent the closure of processing. The sub-process closes with the availability of data files containing data ready for secondary processing and dissemination preparation and adhering to data protection rules.

VII. Prepare for dissemination



The goal of the process phase is to prepare – based on processed data – the dissemination products considering the outputs specified in sub-process II.1. The dissemination products coming about in this process phase will undergo a deep verification and thus will be ready for publication. The process phase contains all activities which will facilitate internal and external users' comprehension regarding the products.

The process phase includes:

- Producing first, draft outputs of the data production process;
- In case of applying seasonal adjustment the production of adjusted data;
- Macro-level validation of produced information;
- Applying statistical disclosure control measures for producing final dissemination materials;
- Interpretation, textual analysis, evaluation of results necessary for final dissemination outputs;
- Production of descriptive, information purpose-serving metadata.

The final dissemination content is created at the end of the process phase.

VII.1. Prepare draft outputs

The scope of the sub-process is to create the first, raw outputs of data production from the available data, as designed in sub-process II.1. Besides it contains the production of all auxiliary information (e. g. indices, trends, quality-related characteristics) necessary for evaluating raw outputs. Data loading into data repositories takes place in this sub-process, as well as the preparation of loading into databases serving dissemination. The sub-process closes with the production of draft outputs and of the auxiliary information needed for their evaluation.

VII.2. Produce seasonally adjusted outputs

The scope of the sub-process is the production of seasonally adjusted time series complying with fixed parameters, according to the methods set forth in sub-process II.6. We include here the methodological verification of seasonally adjusted time series, and if need be, adjustment of models based on time series. The sub-process closes with the production of accepted, seasonally adjusted data.

VII.3. Validate on macro level

The scope of the sub-process is to examine whether the available aggregated data comply with the macro-level validation aspects specified in sub-process II.6. We examine in the sub-process if based on information relating to the given statistics (e. g. mirror statistics, relevant information of other statistics) the outputs comply with the validation criteria. Macro-validation activity typically contains the following:

- Examination of coverage ratio and response rate;
- Comparison of statistical data with data of previous periods, time series analysis;
- Comparison of statistical data with data originating from other data sources;
- Highlighting inconsistencies with other statistics;
- Collation of statistical data with expected values.

The sub-process closes with the gathering of revealed errors, their methods of correction, decision regarding return to previous process phases (data preparation, processing).

VII.4. Apply statistical disclosure control

The goal of the sub-process is to carry out the task of statistical disclosure verification and actual protection of the produced statistical outputs according to the methods set forth in sub-process II.6. (tabular data protection, producing anonymised microdata, preparation of data to be used in safe environment, statistical disclosure control of public use files) as well as the output checking of research results produced in the safe environment. The sub-process closes with the production of outputs provided with statistical disclosure control and with concluding the verification process of research results.

VII.5. Analyse

The scope of the sub-process is to interpret the outputs of the data production process, their textual analysis, their graphic presentation, their preparation for publication according to dissemination criteria and different needs of different dissemination channels. One task of the sub-process is to evaluate (textually describe) – by utilising information on the data production process – the data production process, to establish in what measure outputs comply with the original expectations. Part of the sub-process is the quality evaluation of the data production process and the resulting products. The sub-process closes with the production of analyses regarding the outputs of the data production process, statistical information, descriptive metadata and the quality report.

VII.6. Finalise outputs

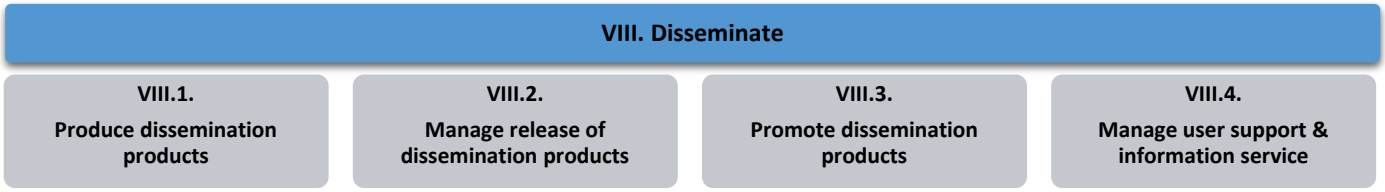
The goal of the sub-process is to determine if statistical data and the linked metadata match the design, the set goals, if they fulfil quality requirements, in consequence if they are suited for use.

It includes:

- Consistency check of final outputs produced;
- Definition of publishability level;
- Finalising documentation for internal use;
- Approval of process- and product quality reports;
- Final content coordination of dissemination products with statistics departments;
- Content approval of dissemination products;
- Finalisation of approved data in data repositories.

The sub-process ends with finalising the content of the dissemination products and related analyses and descriptive documentation (statistical data and metadata).

VIII. Disseminate



We transfer statistical results (data) – through corresponding dissemination channels and communication means – to the user, by adhering to the proper data protection regulations of the different dissemination channels.

Characteristic activities:

- Editing of dissemination products (analyses, data compilations, graphs);
- Loading the predefined tables of the dissemination data into webpages, web-applications, dissemination database;
- Completing outgoing data transmissions, meaning handing over data to internal and external partners;
- Promotion-related activities of dissemination products;
- Providing users with information service activities.

VIII.1. Produce dissemination products

The scope of the sub-process is to prepare the release and editing of the finalised dissemination products as established in sub-process II.1. and finalised in sub-process VII.6 based on data ready for dissemination. Depending on dissemination product type loading of data into display interface, web-applications, dissemination database, editing and visualisations of online and printed materials (e. g. first releases, analyses, yearbooks, media materials, communiques, maps, data visualisations) and preparation of datasets for data access channels provided for correspondingly servicing data requests. The sub-process closes with actual preparation of dissemination products, and providing their accessibility (data to be accessed).

VIII.2. Manage release of dissemination products

The goal of the sub-process is to provide accessibility for dissemination products, their actual release; including time of release. It includes preliminary access of publications under embargo (Press Room, pre-release access according to prevailing internal regulations). In this context dissemination products should be handed over to partners, users requesting data, actual data access should be provided by performing tasks related to dissemination channels (by content service). The sub-process closes by providing user access to dissemination products (released data).

VIII.3. Promote dissemination products

The scope of the sub-process is to promote the dissemination products prepared during the data production process and to make them available through official services for an extensive range of users. It includes the education of certain user groups (e. g. within user forums), their brief guidance. The promotion of dissemination products should rely on the widespread and effective resource system of the media and marketing building it on the Office's homepage as the main dissemination and communication tool. The sub-process closes with the completion of promotional activities regarding dissemination products, where the presence in the social media – besides communicating with the press – is an important component.

VIII.4. Manage user support & information service

The scope of the sub-process is to register users' data- and information demand and their timely and qualitative fulfilment. This task is to be met by the activity of the information provider organisational units and by professional support offered by specialists. In this sub-process decisions must be taken regarding fulfilling certain user demands (e. g. provided through accessible data, new product to be prepared based on data, or the demand cannot be served). The sub-process closes with serving user demands, by customer satisfaction measurement, and by feedback into next periods' planning processes.

Comprehensive areas of statistical data production

Quality management (A.)

A. Quality management

The scope of quality management is to create and operate a framework for ensuring outstanding quality of statistical products, statistical data production processes, and for the organisational framework supporting these. The final goal of all these is the maximal service of users' needs, as we define quality of statistical products based on users' needs: "Quality is the totality of the unit's (product or service) characteristics which influence the capacity to satisfy determined and expected needs." (MSZ EN ISO 8402 – 1996). The products are the result of the statistical data production process, in consequence the statistical data production process and the process quality have a great impact on the product quality. The quality guideline describes our expectations regarding the quality of the individual process phases.

Quality could refer to the organisation, processes and products. The quality management framework builds on the logic of one tool of continuous development, the so-called PDCA cycle[4]. According to the PDCA cycle we set up a plan referring to quality, and define goals based on expectations in order to achieve an excellent product- and/or process quality (e. g. laws, standards, minimum-criteria for product quality components, user-specific demands, quality guidelines). Then data production process follows; we document and measure in between (e. g. quality reports, quality indicators, customer satisfaction measurements). We compare the plans with the documentation and evaluate the adherence to the plans (e. g. self-assessment, internal-external audit, user-professional forum). If necessary, based on evaluation, we define development measures, which will be integrated in the next cycle's expectations.

Statistical data management (B.)

B. Statistical data management

Statistical data handling is a regulated field including all process phases of the statistical data production process. Uniform principles are applied when loading into databases, for name conventions, database tables' (files) system, for maintenance, access rights, in order to obtain concise statistical data assets. Data handling has to ensure efficiency of access and sustained data consistency. Statistical data assets can be identified and traced partly by metadatabases partly by the catalogue of applied software (data dictionaries). Preparation of data handling of a given data production process takes place in sub-process III.1. by preparing the physical plan of the data production process and by creating the database tables. Data handling and access is a comprehensive area concerning every step of process phases IV–VIII.

[4] PDCA: plan–do–check–act.

Applying data protection principles is an important aspect in planning and operating data storage, handling, access. This is very important during personal data handling where different rules apply for handling identifiers, statistical disclosure control, ensuring identifiers 'characteristic criteria and identifiers' connection from those prevailing in the case of handling other microdata and aggregated data (e. g. natural and artificial identifiers' separated storage, data integration rules etc.).

Archiving is part of the unified statistical data production process. Archiving order is determined by archiving policies and the unified archiving regulations built on it, which present suggestions regarding the archiving of different types of product results (depending on frequency, processing grade) following the online access period. Out of the product results the input of the data production process, the data preparation result (i.e. processing phase input) and the finalised data resulting from processing (microdata and aggregated data) – eventually the data ready for dissemination- have to be archived. Archiving includes the storage of data and of the information, metadata, documentations and programs necessary for their interpretation. The archiving design linked to the given process has to be created in sub-process II.6. taking into account the data, metadata format, their availability, the archive loading process, the archived data's cataloguing, the retrieval methods from archive. Archiving itself is linked as such to different process phases defined during design. At the closing of the process or after a predetermined period the verification of archiving has to be performed as well.

Metadata management (C.)

C. Metadata management

The goal of metadata-management as a comprehensive field is to ensure the availability of corresponding metadata during design and execution of the data production process. Metadata management includes the handling of all three types of metadata (descriptive, structural, administrative). Metadata appear in all process phases irrespective of whether they appeared in that process phase or have been utilised from the previous process phase. Metadata management represents the design, production and maintenance of descriptive, structural and administrative metadata, which offer parameters for description and management of process phases and sub-processes.

The process includes the unified regulation of metadata handling. Necessary metadata appear at the same time as the concerned sub-processes, they are linked to these sub-processes and are utilised in the further phases. The process is regarded as complete by providing the corresponding quality metadata for data production and by ensuring regulators referring to metadata management.

Annex – Concepts

The annex contains the conceptual definitions of terms used in the Hungarian GSBPM to create and use a common language. The identification number of any concept already included in the metadatabase was given behind the concept.

- Process phase of data preparation: The fifth process phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM) during which we encrypt, verify and improve the collected and received data to get approved ready-to-process data.
- Data preparation: recording, checking and correcting data collected for statistical purposes or received from other institutions.
- Survey instrument: all technical instruments necessary for carrying out surveys, including edited questionnaires (electronic or paper based questionnaires) and interviewing devices (e.g. laptop, PDA, etc.).
- Process phase of survey implementation: The fourth process phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM), during which data characteristic of the survey population are collected according to the survey method for the reference periods of the survey.
- Survey proposal: As a starting document of the survey design it is an advisory material to start and modify surveys detailing the purpose and content of the new or modified survey and the mode, output and cost of implementing the survey.
- Survey design: As a planning document it defines the content, method, and means of data recording and specifies how to check, process and disseminate collected data.
- Data source: The source of data needed to carry out the statistical survey. Multiple data sources can be used to gather data required for the statistical survey.
- Data collection organisation: After designing the survey, it is the implementation phase of the statistical data production process to collect data including the following tasks: defining survey frames (reference scope), sampling, defining the scope of data suppliers, preparing and delivering forms to data suppliers / interviewers, visiting addresses, registering the receipt of questionnaires, urging the submission of missing questionnaires, registering causes of deficiency, organising the work of participants in data collections and monitoring the organization of data collection processes. Collected data are forwarded by the data collection organisation to the data preparation phase (next phase of the data production process) and deficiency information to the data processing phase.
- Data integration: The process of linking data from two or more sources to generate a statistical output.
- Statistical data domain: Characteristics observed by statistical surveys that relate to a given topic and level of observation within the same logical unit.
- Data labelling: During data preparation and processing, data can be labelled to generate output that may be important information (e.g. corrected, replaced, outlier or protected data) during further processing and quality evaluation. Both rows (record-level labelling) and cells (item-level labelling) can be used in the data table.

- Scope of data suppliers: Scope of the data supplier is the set of entities of the frame population assigned for data reporting from which data can be retrieved for the investigated population (statistical and observation units).
- Administrative data source: Secondary data source, where the collection of data and the keeping of the record is prescribed by law for the owner of the administrative data source.
- Anonymization: A procedure by which the statistical unit concerned (typically a natural person or a business entity) becomes no longer identifiable.
- Archiving: A well-managed, regulated activity and process whereby the final outputs generated during the statistical data production process can be stored in a predetermined, typically long-retrievable and recoverable manner, ensuring their safety in a location physically separated from the primary source of output products.
- Encryption: Handling personal information in a way that makes it impossible to identify the specific natural person whom the personal data relates to without using any further information, provided that such additional information is stored separately in such a way that no personal data can be linked to any identified or identifiable natural person, which is guaranteed by technical and organisational measures.
- Estimation: Estimation is concerned with inference about the numerical value of unknown population values from incomplete data such as a sample.
- Target population: The target population is the population we wish to study, that is, the set of elements about which estimates are required.
- Editing: Statistical activity to detect data and data connection defects, to investigate the credibility of data, and to correct errors.
- Primary data source: a data source from which the characteristics of a population described by the statistical survey related to a specific period or date may be determined wholly or partially through statistical data collection.
- Hungarian GSBPM comprehensive area: A general activity covering all process phases, including quality management, statistical data management and metadata management
- Hungarian GSBPM process phase: One of the main steps or activities that constitute the process of statistical data production. The process phases are as follows: specify needs; design; build; collect; prepare for production; produce; prepare of dissemination, disseminate.
- Hungarian GSBPM subprocess: A subprocess within the process phase that produces a specific output that is different from the outputs of other subprocesses.
- Hungarian Generic Statistical Business Process Model (Hungarian GSBPM): The Hungarian adaptation of the GSBPM (Generic Statistical Business Process Model) created by the UN Economic Commission for Europe (UNECE), the process matrix for the entire statistical production process.
- Reception of data: Comparing the scope of data suppliers and the actually received questionnaires. In a narrower sense, the registration of incoming questionnaires. In a broader sense, reception of data includes contacting and urging those data suppliers who did not complete and return the questionnaire sent to them as well as determining and registering why they failed to return the questionnaire.

- Process phase of Build: The third phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM), in which the means needed for surveying and producing statistical data (e.g. questionnaires, data transfer plan and application systems) are completed and the survey process becomes operational after their testing and documentation.
- Process phase of Produce: The sixth process phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM), during which we produce ready-to-read data from ready-to-process data through linking, replacing, estimating and aggregating data as well as generating new indicators and new statistical units.
- Survey frame: Survey frame is the set of survey population units together with their attributes referring to a given reference period.
- Survey population: Survey population is the population for which information during the survey process can be obtained.
- Correspondence table: connection and matching between the elements of two nomenclatures.
- Error calculation: calculation of sampling and non-sampling errors.
- Process phase of Identify needs: The first phase of the Hungarian GSBPM, during which an advisory survey proposal is prepared based on the analysis of external and internal needs.
- Imputation: Imputation is a procedure for entering a value for a specific data item where the response is missing or it is unusable.
- Coding: A technical procedure for converting verbal information into numbers or other symbols which can be more easily counted and tabulated.
- Macrodata: Aggregate data generated from record-level (individual) data as a result of statistical data processing, or data collected at a higher than elementary level.
- Macro-validation: Detection of errors in statistical data by checking aggregated data, and by checking the entire dataset or a subset based on statistical methods.
- Secondary data source: a data source from which the characteristics of a population described by the statistical survey related to a specific period or date may be determined wholly or partially through statistical data transmission. Two types of the collection from secondary data sources can be distinguished: data transmissions from administrative data sources (administrative data transmissions) and transmissions from other secondary data sources.
- Metadata: Metadata is data that defines and describes other data.
- Meso-validation: A micro-validation to verify that statistical data is related or not related to data belonging to other statistical themes and to detect their errors.
- Microdata: A set of records that contains data from observation units. The primary source of aggregate data production.
- Micro-validation: examination of basic statistical data or their connections to detect data errors.
- Quality assurance plan: A plan drawn up as part of a survey plan including process standards, formal obligations (such as laws and internal rules) and the system of quality control actions to eliminate, monitor and evaluate sampling and non-sampling errors as well as to manage the sub-processes of the statistical data production process.
- Quality requirement: Specific, measurable requirements and expectations regarding the statistical product, statistical data production process or the entire statistical organisation, on

the basis of which the quality of the statistical product, the statistical production process or the statistical organisation can be judged.

- Sample: The sample is the set of sampling units selected from the sampling frame.
- Sampling frame: Sampling frame is an information set for the survey population or the attributes of stratification being used as a basis for sample selection and in subsequent estimation procedures.
- Sample design: The term 'Sample design' is used in a clearly defined sense, with reference to a given frame, as the set of rules or specifications concerning the sample selection process. These specifications describe the sample selection process theoretically, nevertheless in an unequivocal manner. The term 'sampling plan' may be restricted to clearly describe all steps needed practically in the sample selection process. The term 'sample design' usually covers the description of the estimation method as well.
- Measure: A statistical measure measures a periodical social or economic phenomenon.
- Group of measures: A statistical group of measure that measure a periodical social or economic phenomenon.
- Value of measure: In case of a particular statistical unit (a group of statistical units) the value of a statistical measure for a given period.
- Variety of measure: A statistical measure measures a periodical social or economic phenomenon in a modified detail.
- Nomenclature: A nomenclature is a set of groups which are assigned to one or more variables of a statistical population. A nomenclature is a system of names for things.
- Classification: A classification is a grouping of the population's units which consists of exhaustive and mutually exclusive classes and each member of a population can only be allocated to one class without duplication or omission.
- Output: All products that are generated during data collections including statistical data from data collections, statistical tables, microdata, analyses, database tables, publications and other information products.
- Paradata: Typical data describing the process, circumstances and results of statistical recording.
- Sample survey: A test of the suitability of steps and means planned during a sample design, which provides information as follows: clarity of questions and filling instructions, feasibility of planned submission deadlines, length of the interview made by an interviewer and answerability of questions (do data suppliers have the information needed to answer the questions). In case of a sampling based data collection, first of all it is worth examining whether the probability conditions are met.
- Snapshot of register: Snapshot of a register is its frozen state on a given date. Instead of a register, snapshots are used for statistical processing because, unlike register units (that can be updated frequently), population units and their attributes must be constant during data collection and statistical processing.
- Representative statistical survey: We observe an appropriately selected sub-population instead of the total (target) population of objects or their classes and extrapolate our results to the total target population.

Types of representative survey methods by selection and sampling procedures:
  – random sampling (simple random, single-stage, multi-stage, stratified);
  – non-random sampling;
  – systematic sampling.

- Statistical data: Statistical data is a result of statistical observations and further statistical operations concerning characteristics of units in the concrete universe.
- Statistical survey: collection of characteristics on the society, economy or environment, referred to a specific period or date of a population, for statistical purposes, using various data sources.
- Statistical data collection: collection of characteristics of a population described by the statistical survey related to a specific period or date by means of interviewing data providers or direct observation.
- Statistical disclosure control: A set of methods for modifying statistical datasets that minimize disclosure risk as much as possible. The most important purpose of protecting data files against statistical disclosure is to prevent individual data from being identified.
- Statistical concept: A unit of thought with independent meaning related to the subject as well as to the production and operation system of statistics.
- Statistical measure: it represents individual numeric variable values (e.g. mean, mode, total, index, etc.) that characterize statistical units in a given group.
- Statistical nomenclature: List of groups formed according to one or more criteria from elements of a statistical population. A systematic inventory of names and denominations.
- Statistical register: A statistical register maintained by members of the National Statistical Service for the purpose of supporting and standardizing the statistical data production process, from which the current and historical state of the surveys are available including the cause, scope and source of changes, and which stores structured data on the individuals (units) of the population in a database.
- Statistical variable: Specific occurrence of a variety of measure in a dataset.
- Urging: notifying data suppliers (respondents) about the delay before (proactive reminders) and after the deadline for submitting data in the process of data collection and organization.
- Statistical domain: A statistical activity that has common characteristics with respect to concepts and methodologies for data collection, manipulation and transformation.
- Concept related to a statistical domain: statistical display and interpretation of social, economic and environmental concepts observed in the context of specific statistical subject matter domains.
- Personal data: Any information relating to an identified or identifiable natural person (person concerned).
- Seasonal adjustment: The process of filtering out seasonal and calendar effects from original data series. The seasonal adjustment process also includes the treatment of outliers and the production of such trends that are free of accidental factors.
- Process phase of Prepare for dissemination: The seventh phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM) that produces data for information dissemination purposes from processed data including checking, protection against disclosure, evaluation and documentation.

- Process phase of Disseminate: The eighth process phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM) is a complex system for publishing information generated during the statistical data production process including measuring and feeding back of user needs and satisfaction.
- Full statistical number: The statistical number of the survey frame, i.e. the status of the common underlying register determining the survey frames at a given point in time.
- Census: Statistical data collection covering all elements of the observation (all survey objects and their classes), all objects constituting the target or total population, however in practice, the frame population is surveyed, that is the specific survey frame. A survey that focuses on the full range of the population to be observed.
- Process phase of Design: The second phase of the Hungarian Generic Statistical Business Process Model (Hungarian GSBPM) detailing the survey plan from implementing the survey to disseminating information.
- Reference period: Reference period is a time interval or a date to which the observed attribute (variable, indicator, measure) refers.