



**Part II: GUIDELINES FOR QUALITY ASSESSMENT
IN THE NATIONAL STATISTICAL INSTITUTE**

Sofia

TABLE OF CONTENTS

1	INTRODUCTION	3
2	DESCRIPTION OF THE STATISTICAL PROCESS	3
3	PROCESSES QUALITY INDICATORS	4
4	SPECIFIC FEATURES OF THE QUALITY COMPONENTS BY TYPES OF SURVEYS	5
5	CONCLUSION	18
6	ANNEXES	19
	ANNEX 1: GENERIC STATISTICAL BUSINESS PROCESS MODEL (GSBPM)	20
	ANNEX 2: LIST OF PROCESS QUALITY INDICATORS AND DOCUMENTS	21
	ANNEX 3: STRUCTURE AND RECOMMENDATIONS FOR PREPARING QUALITY REPORT ACCORDING TO THE ESS STANDARD FOR QUALITY REPORTS, 2009	28
7	REFERENCES	36

1 INTRODUCTION

This document – Part II: “Guidelines for quality assessment in the National Statistical Institute (NSI)”, completes the following methodological documents on quality developed within the framework of project BG 0203.12 under 2002 National Phare Programme: “Compendium of methodology for quality assessment of statistical data in Bulgaria” and “Instructions for implementation of the quality reports of statistical surveys”.

The purpose of this document is to complete the above-mentioned ones, adding some core elements related to the quality of the European statistical information that occurred as of 2003 onwards. The new Eurostat methodological papers and materials on quality, as well as the good practices of other countries, have been taken into account.

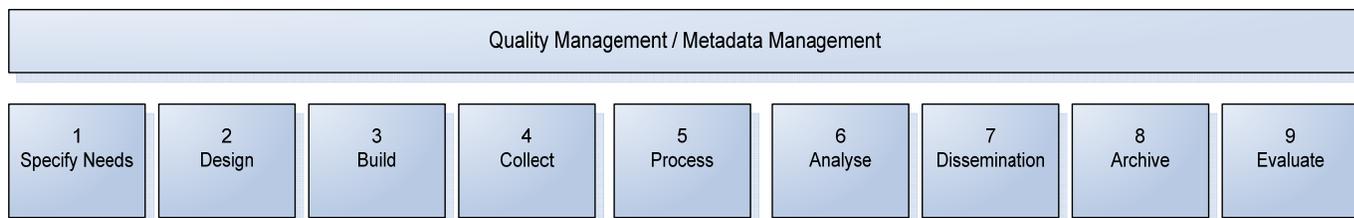
Part II: “Guidelines for quality assessment in the National Statistical Institute” contains description of the statistical processes, processes quality indicators, special features of the quality components by type of survey and structure and recommendations for quality report. In the process of describing the special features of quality components, measuring of quality components has been considered by focusing on the component “Accuracy”.

This paper has been elaborated by the NSI as a result of participation in the project “Quality Assurance in the National Statistical System”, funded by the Statistical Office of the European Communities (Eurostat) and the NSI of Bulgaria.

2 DESCRIPTION OF THE STATISTICAL PROCESS

The Generic Statistical Business Process Model¹ is a group of processes necessary for the production of statistical information, divided by sub-processes in nine phases: specify needs, design, build, collect, process, analyse, dissemination, archive and evaluate. The aggregate of processes and sub-processes creates the architecture of the statistical activity which includes development, production and dissemination of statistical information.

Figure 1: Phases of Generic Statistical Business Process Model



Direct impact on the quality of the statistical surveys has each separate consecutive phase of the statistical process.

¹ Generic Statistical Business Process Model (GSBPM), approved by the METIS Steering Group, is available at: <http://www1.unec.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

Quality assessment of the statistical information is made in accordance with the output quality components, laid down in the European Statistics Code of Practice, namely: Relevance, Accuracy, Timeliness and Punctuality, Accessibility and Clarity, Coherence and Comparability (for additional information please refer to “Guidelines on quality criteria in the National Statistical System”).

In assessing each output quality component, the sub-processes for development, production and dissemination of statistical information are taken into account. Special attention is paid to the quality of information provided by other administrative sources being the main source of information for some surveys, or is used for elaborating statistical characteristics, references, as well as for completing the information available in the registers and information systems maintained by NSI. It is necessary to study the separate phases of the process and to define the main variables which have an impact on the quality and effectiveness thus improving the statistical process. Supervision and improvement of the data quality of each statistical process are of major importance for reaching continuous improvement of quality.

3 PROCESSES QUALITY INDICATORS

Quality indicators are widely used instruments for measurement of quality of the statistical information. They are an important tool for self assessment and audits on quality. By using the quality indicators, statistics experts or users could assess the quality of different surveys or the quality of one and the same statistical information but for different periods. When quality indicators are used for informing users, it is appropriate quality characteristics to be presented which will help in correctly interpreting data or aggregating basic effects on the statistical information usage. Quantitative indicators serve data producers in supervising and assessing the processes over time and finding out the sources of errors in order to improve the processes. Analysis of errors requires availability of comparable information, both for the coverage (of the processes) and the concrete classification of errors used.

The statistical practice has traditional methods for measuring of some indicators on processes, e.g. non-response rate, interview errors, expenditure, time use for different processes.

Quality assessment of statistical processes is currently carried out at NSI, as well as calculation of different indicators on the quality of products is done, but without applying a systematic approach. The information is used for different purposes, for example for determining the non-response unit rates, for monitoring and control of the stochastic error in sample surveys. For presenting data on quality, different forms are used, for example the indicators included in the quality reports, information on the activity of different sub-processes, in the report on the implementation of the National Statistical Programme and the NSI’s activity. Different criteria are used in determining the indicators, for example those according to the Eurostat standard indicators, as well as criteria already adopted in the national practice, etc. Indicators are calculated with different periodicity – in accordance with the periodicity of the surveys, the Eurostat requirements or one time only upon publishing the relevant information, etc. Information on quality of the processes is collected and stored in different places – by the experts responsible for the specific survey, by the experts responsible for the dissemination, as well as at different organizational levels – Regional Statistical Offices (RSOs) and the Head Office. For the time being data on quality of the processes are used mainly in designing standard indicators on the quality of products, in carrying out self assessment on the quality of different surveys or as an information source for defining different sub-processes.

This paper contains description of the statistical process phases and the relevant sub-processes, as well as List of quality processes indicators by sub-processes (*please refer to Annex 1 and Annex 2*). The motives to include the description in the list are:

- Priority setting of the process approach in the development, production and dissemination of statistical information regarding the Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: a vision for the next decade²;
- Implementation of complex assessment on quality of the statistical information;
- Improving the communication with users.

The logical way of the phases of the statistical process and sub-processes is related to quality assessment based on combination of two approaches by using indicators (quantitative and qualitative) and evidence based approach.

In this way conditions for realization of complex assessment and monitoring of quality are created by using information on indicators and availability of documentation checked. Basic documents are included that characterize the institutional environment in which the statistical processes are carried out.

In designing the list of quality indicators by sub-processes, a differentiated approach is applied. For each of the nine phases, the sub-processes are described and an attempt has been made to include relative quality indicators or the implementation itself, as well as documents. It should be noted that only the basic indicators/documents are pointed out. The list is still open for further updating.

4 SPECIFIC FEATURES OF THE QUALITY COMPONENTS BY TYPES OF SURVEYS

The National Statistical Institute carries out six main types of statistical surveys defined in the ESS Standard for quality reports, 2009. The six main types of statistical surveys are:

- Sample surveys;
- Censuses (exhaustive surveys);
- Surveys based on administrative sources;
- Surveys that combine data from different types of sources, so called statistical processes involving multiple data sources;
- Surveys in which prices or other economic indices are calculated;
- Statistical compilations.

Because of the specificity of some of the surveys carried out by the NSI it is difficult to refer those only to one of the six types of statistical surveys. For example, the statistical survey of consumer prices can not be classified to one of the specified types of surveys because it is:

- sample survey, most of the samples apply so called non-probability sampling techniques (cut-off);
- combined survey and uses data from different types of information sources in prices calculation as well as weights construction (administrative sources are also used in some cases);

² [COM \(2009\) 404 – regarding a vision for the next decade and implementation of method for integrated preparation of the European statistical information.](#)

- price indices are calculated.

Different process specificities should be taken into account in the quality assessment of these surveys.

For all types of surveys quality assessments by separate components are developed. In the different types of surveys unified approaches are used for measuring the specific components. Quality assurance depends on several basic factors: up-to-date methodology applied, sample representativeness, organization of the statistical process, scopes and effectiveness of the control, information and communication technology application, qualification and skills of the personnel. It should be taken into account that the specific features of the separate survey groups have an impact on the relevance and significance of the different quality indicators by components.

In addition, major aspect for improving the quality of statistical process is the activity carried out by all 28 Regional Statistical Offices (RSOs) being the main actors in the implementation of NSI's functions and tasks for the production and dissemination of statistical information at regional level. They are directed mainly to: acquainting respondents with the purposes and tasks of the statistical surveys and motivating them to take part in the process of data provision; providing guarantee that the secrecy of individual data on physical and legal entities will be kept; providing clarifications on the methodology and instructions laid down in the tools of statistical surveys; training and assisting respondents on the spot or at the premises of the RSO; training the household members how to fill in the questionnaires on the household budget consumption; sending letters with methodological explanatory notes to the responding units; reflecting up-to-date changes in the content and calculation of separate indicators; consulting them on different issues as regards survey methodology and organization.

Relevance in measuring the “Accuracy” component by types of surveys is presented in Table 1. The aim is to determine in a comparable dimension the relation and meaning of the measure “Accuracy” in the separate surveys. Examples are presented in order to illustrate the activities on quality assessment by phases and by sub-processes, as well as guidelines for improvement are provided.

Table 1: Some specifics of accuracy component, broken down by type of statistics

		Surveys		Using Administrative Source(s)	Involving Multiple Data Sources	Indexes	Statistical Compilation	Registers
		Sample	Census					
	Sampling effects	XXX			X	X		
Non-sampling effects	Evaluation of data sources	X	X	XXX	XXX	X	XXX	XXX
	Coverage	X	X	XXX	XXX	XXX	XX	XXX
	Non-response	XXX	XXX	X	X	X		
	Measurement errors	XXX	XXX		X			
	Processing errors	XXX	XXX	X	X	X	XXX	XXX

XXX - very high relevance
 XX - high
 X - medium

It is widely known that “Accuracy” is accepted as a traditional aspect when assessing the quality of statistical surveys, but in terms of applying a complex assessment of quality it is considered at equal level with the other components.

Naturally, sampling errors appear to the largest extent in sample surveys and to an average one in surveys that are carried out by combining data from different sources and indices (to the level on which sample surveys could be used). They are related to assessment errors and model selected.

Non-sampling errors are due to:

- quality of the administrative data sources and the coverage. They are most important for surveys based on administrative sources and in combination of data from different sources, statistical compilations and registers;
- non-response, measuring and processing errors and they are the most typical ones as regards sample and exhaustive surveys.

Sample surveys

The methodology of these surveys includes obtaining of information about the general population of surveyed units by measuring a part of them (a sample or a representative population). Statistical characteristics are completed by assessment of the size of the sampling error that contains the differences due to the fact of surveying only a part of the general population.

The most significant aspect for assessing “Accuracy” is the calculation of sampling errors, non-response errors, measurement and processing errors. In order to obtain a reliable picture of surveyed phenomena and processes, the samples should be updated periodically – for example, inclusion of new survey units such as enterprises, households, buildings or exclusion of no more existing ones should be reflected. This requires updated maintenance of the relevant registers and preparation of up-to-date lists of statistical units that will be included in the surveys.

EXAMPLE 1

The stochastic accuracy of the results in the **Labour Force Survey** is estimated using coefficients of variation under the SPSS software. The coefficients of variation are calculated for main indicators – number of employed persons, part-time employed persons, number of unemployed persons, unemployment rate, and average number of hours actively worked by one employee. The coefficients of variation for some of these indicators are calculated also by statistical regions and districts. Calculation of the confidence intervals is done for employed persons, unemployed persons and unemployment rate (based on the probability approach). These intervals indicate with certain guarantee probability the borders of the population parameter.

The non-sampling errors are related to presence of non-responded households (average for 2009 it was about 16%, of which 11% due to no contact made with the households and 5% due to refusals), measurement errors and processing errors. In case of non-response, imputation is applied only for specific variables (economic activity, occupation, unemployment period) and the percentage of imputed values is very small. Statistical methods as method of historical imputation, “hot-deck” imputation are applied.

Errors in the process of registering are made by the respondent or by the interviewer (the share of indirect interviews is about 40%). Many logical checks of the software used for data entry are made, and at the same time discrepancies are corrected. Additional logical compatibility of data in the aggregated files is carried out. Sometimes errors are made in codification of the economic activity of the local economic units, the occupation, and the field of education. Overcoverage error is calculated as share of uninhabited dwellings included in the sample, which is about 8%.

Main directions to improve quality of the LFS are: reducing the non-response rate and using methods to decrease the impact on the survey results; improving the procedure for data weighting; improving the selection method; focusing on training and motivating interviewers; gradually implementing new methods for data collection; improving coherence of data with the national accounts.

Censuses

The main specific feature of censuses is that of major importance for their accuracy are the non-response errors, measurement errors and processing errors of primary data.

Taking into account the pattern of carrying out exhaustive surveys, all exhaustive surveys are mandatory and for the greater part of them the provision of primary data is the responsibility of the Regional Statistical Offices (RSOs). The RSOs organize the work on collecting individual data from the respondents and do the quality control in the process of data collection. At the same time the quality of sub-processes in designing and implementing the survey has an impact on the process of data collection. For example, very important for the data collection process is the quality of the statistical tool. The statistical questionnaire is the key element of the survey tool. It is a tool for identifying the information needs by containing relevant concepts and definitions. In addition, it is a tool for standardization and control on the surveys. Currently the number of electronic questionnaires provided by enterprises is increasing. This fact has a positive effect on the response rate. In principle, the response rate for the big and medium enterprises is relatively high, but there are some difficulties regarding the data provision from the small entities.

Population and agricultural censuses are such surveys that have concrete requirements for quality. For example, according to the draft Regulation of the Commission for implementing Regulation (EC) No 763/2008 of the European Parliament and of the Council of 9 July 2008 concerning the Population and Housing Census as regards conditions, structure of the quality report and the technical format for data transmission, it is envisaged the Member States to provide by 31 March 2014 to the Commission (Eurostat) a report on the quality of the submitted data and metadata with main focus on the application of the quality criteria.

Surveys based on administrative sources

Usage of administrative sources in the statistical process is increasing continuously being the source of primary data and additional resource for realization of the tasks in the separate phases and the relevant sub-processes (for example, in the selection of samples, revision, validation and editing of data, imputation, etc.). This process is also related to reducing the costs for data collection, respondents' burden and the given opportunity to provide detailed information to users at regional level as on this level it is difficult to obtain representative results from the sample surveys. Administrative data sources are registers or information systems established as a result of the administrative work of the relevant institution and they are out of the control of the NSI. Assessment of accuracy is based on the information provided by the administration that collects and manages these data.

The main requirements for their usage for statistical purposes are:

- To be supplemented with detailed metadata;
- To provide information for identification of the surveyed units;
- To be provided in an appropriate format and structure that will allow effective combination of records from other files with administrative data.

To assess the accuracy of surveys based on administrative sources different approaches are applied by taking into account mainly the errors in the data source and the coverage. The main difficulty in assessing outside sources is the difference in definitions stated in the administrative sources with the ones used for statistical purposes. It often happens that a certain administrative source contains enough information for a greater part of the surveyed units, but at the same time gaps exist.

Quality assessment of the administrative sources could be completed by assessing the cooperation with the relevant administrations maintaining administrative data with which NSI has signed agreements for exchange of information. In this connection assessment should be made on the following³: length of usage of a certain administrative source; conditions for data exchange; clarity of the requirements for confidentiality; completeness of the relevant metadata; timeliness of the information when changes in the legal basis of the relevant administration were made.

EXAMPLE 2

Surveys on population are based on administrative data from the Unified System for Civil Registration and Administrative Services to Population (USCRASP) – unified administrative documents for registration of birth, death, civil certification of marriage, cancellation of marriage, migration (current address registration) as well as data from the Ministry of Interior on foreigners with permanent residence in the Republic of Bulgaria.

Data accuracy is ensured in accordance with the developed methodology on demographic surveys and processes. Verification of the data files (arrays) is done through the information system “Demography” as follows:

1. Entry formal control including validation checks on the codification used for identification of persons and demographic events; double and incorrect records; completeness of the record and accessibility of the indicated variables;
2. Logical control and comparability checks between different sections – links between the sections for characteristics of persons and demographic events – e.g. link between age and marital status, age and education, age and marriage, etc.;
3. Treatment of errors – errors are corrected by additional specification or referring to the primary document, or to the relevant source maintained by the General Directorate “Civil Registration and Administrative Services”.

Validated and adjusted data are entered into the system for generating tables on the population and demographic events, according to the adopted methodology.

In surveys related to **Foreign Trade** different administrative sources and information systems are used.

Statistical procedures on the completeness and reliability of the data files in the Information system “Foreign trade” is carried out in three phases:

1. Entry formal control including validation checks on the commodity codification used and, when required, supplementary unit quantity for the goods upon availability, double or incorrect records, completeness of the record and accessibility of the indicated variables;
2. Statistical control – compatibility checks between the different sections; ratio between net mass and supplementary unit quantity; price variations of the basic trading goods; relations between partner-country and mode of transport; commodity code and mode of transport; ratio between value and statistical value depending on the terms of supply, etc.;
3. Statistical processing of errors – if additional specification or reference to a primary document is needed when

³ Guidelines for assessing statistical processes based on administrative source(s) with the ESPAC Checklist, Working Group “Quality in Statistics”, 2010

trading with a third country, a claim to the Customs Agency is made for additional checks and further corrections. In case of wrongly filled in Intrastat declarations, the relevant firm is contacted for further clarifications regarding data and circumstances, and when necessary, the firm is asked to provide a new declaration to the Information system Intrastat of the National Revenue Agency (NRA), then NRA submits it to NSI.

After checking and adjusting of primary records the data is loaded in final registers of the Information system “Foreign trade” where on the basis of corresponding tables maintained by the NSI the data of the Combined nomenclature refer to the other classifications – SITC, NACE, etc.

In trading with EU countries, where excluding survey threshold is applied, each month an additional estimation beyond the threshold is made using statistics-mathematical models and data from the Tax declarations for the intra-community supplies and acquisitions. Assessment of non-response of persons with liabilities is carried out. The additional estimates are added to the main file and the obtained data become part of the volume forming commodity flows in reporting the intra-community trading with goods. Data from final registers of the Information system “Foreign trade” is initially loaded in the system internal cubes intended for the NSI users and then published in the system external cubes and accessible via Internet to all users. The data arrays (in defined structure and format) are presented to main users in Bulgaria (Bulgarian national bank, Ministry of economy, energy and tourism, etc.) and abroad (Eurostat, UN, etc.)

On a quarterly basis foreign trade price indices are calculated according to the approved methodology. When the final annual data array is ready, a revision of the quarterly and annual index for the reference year is accomplished.

Quality assessment of administrative data should be done in line with the Eurostat guidelines for assessing statistical processes based on administrative sources⁴, the Strategy for Development of the National Statistical System 2008-2012⁵, as well as with some good practices of EU Member States.

Quality improvement is in close relation with elaboration of a mechanism for monitoring the changes in the national registers and information systems; usage of administrative data for improving the methodology of the statistical surveys; implementation of modern information and communication technologies for data exchange between NSI and the institutions maintaining administrative sources; improving and updating the Register of Statistical Units.

Some EU statistical offices have applied in practice own developments which are of common interest. A standard framework for assessment of administrative data has been developed by Statistics Netherlands by which quality assessment as regards data source, metadata and statistical data is done. They use a special questionnaire to collect information on specific indicators in order to determine the quality of the source and metadata from the quality framework. The Statistical Office of Portugal had carried out an internal quality survey on the administrative sources and the access to administrative data by all statistical units. The information was collected through two questionnaires on:

- Assessment of the institution responsible for the administrative data, its strong and weak points (regarding deadlines for data provision; relationship between NSI and the other administrations, etc.);
- Assessment of the future needs of administrative data. In this way important aspects for improving the access to administrative data and their quality had been identified.

Surveys that combine data from different types of sources

⁴ Guidelines for assessing statistical processes based on administrative source(s) with the ESPAC Checklist, item 4b–12th meeting of Working group “Quality in Statistics”, June 2010

⁵ [Strategy for Development of the National Statistical System of the Republic of Bulgaria, 2008 - 2012](#)

The principle “collect only once primary data and use them many times and for different purposes” becomes leading in the realization of the data collection policy. Combination of data from different sources within the State Administration and other public institutions, aiming at creating an information resource for specific statistical purposes, is a guiding line for producing new products, reducing the statistical data collection and production costs and reducing the respondents’ burden.

EXAMPLE 3

Typical case of a survey based on information from different sources is the **European System of integrated Social Protection Statistics (ESSPROS)**. It is a specific tool for statistical observation of social benefits and their financing for social protection purposes in the EU Member States. ESSPROS is composed of three modules – module of the Core System, module of Pension Beneficiaries and module of Net Social Benefits. Data of the Core System provide a possibility to compare quantitative and qualitative data on receipts, expenditures and types of social benefits by social protection schemes. The statistical unit of ESSPROS is the scheme.

The EU Member States have to submit to Eurostat quality reports on an annual basis according to Annex II: Criteria for measuring quality, item 3.1 of Commission Regulation (EC) No 1322/ 2007 on appropriate formats for data transmission, results to be transmitted and criteria for measuring quality for the ESSPROS core system and the module on pension beneficiaries of 12 November 2007.

Main data source are the administrative sources – total for all countries included in the survey their share is about 84%, the share of surveys and module based is the same – 12% in total. Bulgaria uses 17 sources in total, of which 12 are administrative registers and information systems and 5 by conducting a survey – relatively 70% and 30%.

For assessment of accuracy different approaches are used taking into account mainly errors in the data sources and their coverage, the concepts applied, definitions, classifications which not always are in line with the statistical ones. The influence of NSI in the selection of approaches for collection of data from administrative sources is not always satisfactory for achieving coherence and comparability due to the fact that the relevant administrations follow their own administrative and legal frame. Changes in the legal acts lead to interruption of time series. Assessment of accuracy implies specific approaches to be used, searching for data for small sub-groups and applying different statistical techniques in order to overcome the lack of primary data. In cases of lack of available data, schemes are developed by assessments/estimates through different methods, for example temp or structure for the previous year, combination of information from different sources; through the method of complete merging (if there are no data on receipts and expenditures through some schemes); incomplete probability merging or statistical modeling.

Of major importance for improving the quality of surveys based on information from different sources is the cooperation between the Bodies of the National Statistical System and the institutions of the State Administration that maintain administrative data, as well as other public institutions within the frame of the established coordinating mechanism. Further continuous efforts are needed towards harmonization of the available concepts, definitions, classifications and approaches in the collection of statistical and administrative data.

Surveys in which prices and other economic indices are calculated

Statistical surveys on prices and calculation of indices are carried out with the aim of characterizing the inflation processes. A system has been created to measure the relative changes and dynamics of prices. Prices are surveyed in different sectors and activities by studying the following: the needs of statistical information, the requirements of the socio-economic theory and the proved in practice theoretical background, organizational principles and procedures. They are in line with the EC Regulations in the field of price statistics, as well as with other international standards and recommendations.

Of major significance in formulating the concrete requirements regarding accuracy in surveys on prices is the usage of unified methodological principles and approaches to design the coverage of the two components – prices and weights.

Unified methodological principles allow to measure the change in level of prices in different sectors and activities, as well as to assess their contribution for analyzing the dynamics of the phenomena and processes in the economy.

EXAMPLE 4

For the **Harmonized Index of Consumer Prices (HICP)** “Reliability” is assessed according to “Accuracy” which is related to the scope of errors in designing the sample and the “Representativeness” which is related to lack of fluctuations according to Regulation (EC) No 1334/ 2007, Art. 2a, Item 10.

1. Coverage and samples

To ensure coverage according to the purpose of the survey, four samples are used based on the target (expert) selection.

- Consumer basket

It includes goods and services for the final consumption of the households. The following principles are met: coverage of fixed (permanent) for the year set of goods and services which reflects the average change of level of prices of all goods and services that are typical and are in accordance with the structure of the households’ consumption; in accordance with the definitions in the system of national accounts regarding individual consumption and its coverage; relation to the nomenclature of goods and services on the basis of which international price comparisons are carried out according to the European Comparison Programme; inclusion of goods and services which prices are determined following administrative rules.

- Populated places

Representativeness of the output data on level of prices depends to a great extent of the geographic coverage – populated places and population coverage. It covers 27 district centres in which lives over 46% of the population and where the share of retail sales is over 65%.

- Price observation points (outlets)

Prices are collected in selected observation points – sample of stores, market stalls, restaurants, cafes, etc. The selection is done at regional level by the price collection specialists. The number and the structure of the observation points are determined in a way that can assure the optimum number of prices collected, which are sufficient to represent national prices for any of the observed group of goods and services. The number of observation points is determined proportionally to the population in the selected district centres and to the volume of retail sales in the relevant outlets.

- Collected prices

A sample of prices containing the prices of the observed goods and services in the consumer basket.

2. Weights

HICP weights are constructed according to the requirements of the Commission Regulation (EC) No 2454/97 and of the Council Regulation (EC) No 1688/98. The weights for HICP reflect the structure of the household final consumption expenditures.

3. Specific procedures

- Treatment of missing observations and replacements
- Treatment of quality changes of the observed products

In case of temporary missing price, estimation is done in the 1st and the 2nd month by imputation using the method of matching samples and applying the short-term approach – imputation is done using the chain index.

The following most influential methods for assessing changes in the quality of the observation products is applied: method of direct comparison, implicit quality adjustment methods. The procedures are carried out at national level according to Commission Regulation (EC) No 1334/ 2007, Art. 5.

- Introduction of new and significant goods and services during the annual revisions of the sample of goods and services representatives
- Treatment of price reductions – seasonal sales, discounts, special offers – these process are included in the index when they are available to all potential consumers, according to Commission Regulation (EC) No 2602/ 2000, Art. 2

- Treatment of seasonal products – the fixed weight approach is used for determining of the weights for seasonal items, i.e. during the whole year indices of seasonal products are calculated with fixed weights according to Commission Regulation (EC) No 330/2009

4. Control

Data quality checks and validation work is done both at regional and national level.

Data validation process at regional level is divided into two stages:

- During entering the collected prices into the computer system for processing the information;
- After processing at regional level, the first results are checked and validated by specialists in RSOs and, if necessary, prices are cross-checked in outlets.

Validation of data at national level is done after the first index calculation for the current month. The following prices are checked and validated: extreme price levels, extreme or unusual price levels/changes, missing observations, closed or newly opened outlets, changes in the specifications of the goods/services, seasonal products, etc. There is no automatic rejection of collected prices. Each problematic price is considered individually and modifications are done only if needed on the basis of the relevant information.

Directions for the development of statistics on consumer prices are outlined in the Eurostat Strategy on the multi-purpose statistics of consumer prices. In accordance with the vision for statistical production in the next decade and applying integrated approach in different fields according to the quality requirements, the motives for its elaboration have been outlined, namely: multi-purpose approach, product oriented, efficiency and cost effectiveness, structural approach and quality improvement.

Statistical compilations

Specific types of statistical surveys are the statistical compilations where data from other surveys and administrative sources is used.

Bulgaria's **National Accounts** are designed according to the European system of national and regional accounts (ESA'95). ESA'95 is a harmonized European methodology for elaboration of national accounts, adopted with Regulation (EC) No 2223/96 of the Council of 25 June 1996.

Main macroeconomic indicator of the system of national accounts is the Gross Domestic Product (GDP). NSI produces annual and quarterly data on GDP and its components using three approaches – GDP by Production approach, GDP by Income approach and GDP by Final Expenditure.

Different data sources are used and different approaches are applied in calculation of the GDP data and that implies the results achieved to be balanced and GDP estimates to be adjusted.

Of major importance to assess "Accuracy" of the national accounts is to assure quality assessment and internal coherence of input data (surveys, annual financial reports, administrative sources); coverage errors and exhaustiveness of the data sources used and the processing of input data used to estimate the indicators of the national accounts system.

The data production process includes monitoring and analysis, for example:

- Data validation adjustments aiming at adjusting fluctuations, errors, gaps in data sources;
- Conceptual adjustments aiming at harmonization of definitions in line with the European System of Accounts (ESA'95);

- Additional estimates to assure exhaustiveness of data: that includes additional estimates on the non-observed economy;
- Balancing adjustments – in order to ensure balance between resources and their usage.

The data production process in the field of national accounts has three phases:

- Data collection from the available information sources (statistical surveys, annual financial reports of the economic units and administrative data), the so called input variables;
- Data are calculated at constant prices in order to measure the effect of changes in the volume and changes in price levels;
- Compilation of indicators of the national accounts system.

Input variables are verified as regards their coverage, reference period, indicators, measuring units and coherence (to accounting data, between quarterly and annual data, data from other statistical domains and over time).

There are two main sources for accuracy – the statistical fluctuations and analysis of revisions.

Statistical fluctuation is the measure for the reliability of the assessment made as it shows errors and gaps in the national accounts. Great number of different information sources is used to compile data in the field of national accounts. Due to that fact it is difficult to summarize measures for reliability of aggregated estimates. If for each of all data sources sampling and non-sampling errors are known, the complexity of the process through which GDP is assessed makes it difficult to obtain a common estimate on the accuracy of the group of components. The process of integrating the three approaches for GDP production (GDP by Production approach, GDP by Income approach and GDP by Final Expenditure) in a final and balanced way is presented by Input-Output Supply and Use Tables which provide additional information on the raw data accuracy and their coherence with other sources. As a result more accurate overall GDP assessment is achieved.

In data production process in the field of national accounts there are two types of revisions – current and exceptional. The reasons for carrying out revisions are related to: availability of new primary information, improvement of methodology for assessment of indicators, improvement of basic statistics, change in the definitions and classifications used. Revisions show the level of changes of the national accounts data (quarterly – flash estimates and regular estimates, annual preliminary, annual and annual final data after compilation of Supply-Use tables). The revisions and the reasons for their accomplishment present the accuracy of data as it could be considered that estimates are made in order to obtain real values and that they are based on as much as possible reliable sources.

On the NSI website (www.nsi.bg), chapter “Gross Domestic Product”, the revisions of all quarterly and annual GDP data for the period 1995-2010 are published. The revision comprises of the following elements:

- Additional estimates on informal (non-observed) economy (underground activities, income in kind, tips and shuttle trade);
- Financial Intermediation Services Indirectly Measured (FISIM) by users in accordance with Council Regulation (EC) No 448/98 and Commission Regulation No 1889/2002 for the period 1995-2001. Before the revision in 2010, GDP data on the allocation of Financial Intermediation Services Indirectly Measured was reported on users’ sectors only for 2002 onwards. In order to

obtain comparable data for the whole time series, after the revision the GDP data was presented by the method for the indicator measurement also for 1995 onwards;

- Introducing modified cash-based method in reporting taxes and social contributions in accordance with Regulation No 2516/2000 on taxes and social contributions for the period 1995-2010. Through this method on the basis of cash receivables by types of taxes, estimates are made on the share of taxes and contributions actually paid to the moment of reporting (out of the accrued ones).
- The revised GDP data include both data in nominal values and reconciled data at comparable prices. By finalizing the revision, NSI had provided GDP data, reconciled at prices for 2001. According to the Eurostat Programme for data provision, all countries should send GDP data, reconciled at prices for 2000;
- Since September 2011 a new Eurostat requirement for implementation of national accounts data according to NACE.BG-2008 and reconciled at prices for 2005 was enforced.

On the basis of the GDP revision and the provision of fully comparable annual and quarterly data for the period 1995-2010, the macroeconomists could be able to analyze the GDP dynamics for the whole 15-year period and prepare macroeconomic forecasts.

In addition, the NSI calculated and published seasonally adjusted quarterly GDP data – in total and by components for the period 1997-2010.

Seasonal adjustment of statistical indicators in the NSI is organized in accordance with ESS Guidelines on Seasonal Adjustment⁶. The procedure is accomplished through the developed by Eurostat Demetra software, where the TRAMO/SEATS algorithm is applied. Seasonal adjustment approach is applied for the quarterly data on the GDP components by production approach, income approach, final expenditure approach, data on number of persons employed and number of man-hours worked.

EXAMPLE 5

Quality assurance of the Gross National Income (GNI)

According to Art. 5(1) of the Council Regulation (EC, EURATOM) No 1287/2003 on the harmonization of gross national income at market prices (GNI Regulation) Eurostat shall verify the sources and methods used by Member States to calculate GNI.

Eurostat carries out verification of the procedures and basic statistics used to calculate GNI and verification of the data transmitted in the GNI Questionnaires and the Quality Reports.

According to the recommendations and procedures laid down in Art. 4(2) of the GNI Regulation, Bulgaria provides the Commission (Eurostat) with up-to-date inventory of the procedures and basic statistics used to calculate GNI and its components according to ESA'95 (the GNI Inventories).

The Member States improve and update the description of sources and methods for the GNI estimates – so-called “Inventory” in accordance with the recommendations provided. The Inventory is the main instrument for assessing the comparability, reliability and exhaustiveness of the GNI data. The description of sources and methods for estimation of the GDP and GNI data includes Process Tables that illustrate the results achieved at each single stage of the process of

⁶ ESS Guidelines on Seasonal Adjustment:

http://epp.eurostat.ec.europa.eu/portal/page/portal/product_details/publication?p_product_code=KS-RA-09-006

compilation of national accounts – starting from data sources and ending with publication of national accounts data. In order to ensure comparability, the Member States provide the GNI Inventory and Process Tables according to an agreed structure approved by the GNI Committee. Countries update their GNI Inventories within one year of a major benchmark revision, or, where no benchmark revision policy exists, at least every five years.

Eurostat verifies the description of sources and methods for the GDP and GNI estimates (Inventory) and Process Tables by using GNI Inventory Assessment Questionnaire (GIAQ). The purpose of the GIAQ is to ensure systematic, consistent and fair approach to the assessment of the quality of GNI data.

For the purposes of transparency regarding GNI, and if necessary, information visits are envisaged to be carried out in Member States by the Commission's services and representatives from other Member States. The visits aim at collecting additional information that is necessary to identify possible weaknesses of the GNI Inventory, of the statistical sources and methods used to calculate GNI and its components.

Eurostat presents a mission report to the GNI Committee for a full discussion. A list of action points agreed with the country is annexed to the mission report. The action points may call for possible improvements of methodology; the submission of new information that should be included in the GNI Inventory; the redrafting, clarifying or correcting already existing information in the GNI Inventory and the submission of new information that does not necessarily have to be included in the GNI Inventory. At the end of this analysis, Eurostat provides an assessment report on whether the Member States' figures are appropriate for own resource purposes or whether further corrections and improvements are necessary. The assessment reports are presented to the GNI Committee for a full discussion.

GNI data provided by Member States in the GNI Inventory Assessment Questionnaire and the Member States' Quality Reports are verified by Eurostat.

Member States provide the Commission (Eurostat) with figures on GNI and its components compiled according to the national accounting procedures before 22 September of each year as it is stated by the GNI regulation, Art. 2. Member States provide GNI questionnaire and quality report. The questionnaire containing information on GNI includes previous year information and possible changes in past years data. When Member States provide the GNI questionnaire to the Commission (Eurostat), they transmit a quality report on the GNI data. The Quality Report provides the necessary information which describes and shows data compilation process namely on cases where there have been significant changes to the procedures used, basic statistics used and any changes made to the GNI forecast figures for previous years. The quality report is intended to give a brief and concise update on a Member State's methodology, information sources and results.

The verification of Member States' data by Eurostat is based on the consistency of the data over time and the consistency of the data with otherwise published national account data.

According to the GNI Regulation, the information provided by Member States on the statistical sources, procedures for calculation of GNI and its components is verified by the GNI Committee each year. The GNI Committee adopts a formal opinion on the consistency and authenticity of the Member States' GNI data for own resource purposes with respect to reliability, comparability and exhaustiveness. The basic documents on which the verification is based are clearly stated in the opinion of the GNI Committee. Assessment of the reliability, comparability and exhaustiveness of GNI and its components should be done taking into account the principle of correlation between expenditures and benefits.

A systematic approach is needed to be applied in order to improve the quality of data in the field of national accounts and to assure consistent, exhaustive and timely information. Regular monitoring and publication of quality indicators on national accounts is needed.

In order to assure the quality of national accounts data, the following approaches should be applied in data revision and validation:

- Verification and harmonization of data with external sources, e.g. with data from other surveys or from administrative sources;
- Verification of the internal coherence of data;

- Verification of records of units with the largest share and impact on the elaboration of aggregated estimates and for basic information from the business surveys and other statistical surveys used to compile the national accounts data;
- Quality assurance of statistics in the process of data collection and processing;
- Description of the approach used in seasonal adjustments of the quarterly data.

The main directions in the field of national accounts related to the methodological development and improving the quality of data are outlined in the Strategy for Development of the National Statistical System 2008-2012.

Business Register

Availability of a Business Register for statistical purposes is a very important prerequisite and a necessary tool for a coordinated and qualitative implementation of the separate phases and sub-processes of the statistical process. Its usage is of great significance for assuring quality of the surveys as it provides an integrated system of standard statistical units and unified classifications and nomenclatures necessary for systematizing the survey units and designing the general population. Based on this, the Business Register serves as:

- A tool for identification and single determination of the statistical units forming the general population;
- Information source for the up-to-date state of the statistical units' characteristics and for analysis of the general population. Of major importance for assurance of quality information on statistical units is to regulate which specific characteristics could be changed on the basis of reports with no need of administrative documents, as well as to determine the optimum periodicity for carrying out updates;
- Information source for the interaction and dependency between the different types of statistical units. For carrying out and comparing results obtained from different statistical surveys not only availability of a detailed list of all surveys is needed, it requires also availability of detailed information on the hierarchical links and coordination among them;
- A tool for including and accumulating administrative information for statistical purposes;
- Direct source of statistical data on the following indicators: number of employees, net assets of sales/turnover, expenditures for long-term assets, etc. that are used as basic stratification variables in designing samples;
- A tool for measuring the burden on the statistical units by loading, preservation and analyzing the information on the participation of each registered unit in different statistical surveys.

To assess the “Accuracy” component of the Business Register different approaches are applied taking into account mainly errors in the data sources, the coverage and during the processing. Availability of detailed metadata provides opportunity for monitoring of accuracy.

EXAMPLE 6

Business Demography

Business demography data are based on data from the Business Register maintained by the Bulgarian NSI. The Register includes all active enterprises on the territory of the Republic of Bulgaria. There were no limitations in the selection of enterprises for the relevant years, including no limitations for the number of employees or turnover.

With the adoption of Regulation (EC) No 295/2008 of the European Parliament and of the Council of 11 March 2008 regarding structural business statistics, the legal basis was created and the general framework was designed for the collection, processing, provision and assessment of harmonized statistical data on demography of enterprises in the European Union.

Subject of the survey is the statistical unit of “Enterprise” type as defined in Regulation No 696/93 of the European Parliament and of the Council of 15 March 1993.

The main goal of the study on Business Demography is the evaluation of the number of active enterprises and the number of employees. This goal provides very important information on the enterprises’ length of life by type and size, number and type of dead enterprises, rate of growth of the enterprises, and annual survival coefficients for the interval 2004-2008. These data complement the structural business statistics and form the foundation for weighting and computing other statistics for the enterprises.

In measuring other quality components in different types of surveys, relatively unified approaches are applied. Guidelines for assessment and management of the components “Relevance”, “Timeliness and Punctuality”, “Accessibility and Clarity” and “Coherence and Comparability” are provided in the “Guidelines for quality criteria in the National Statistical System” (*for additional information please refer to item 4.1, item 4.3, item 4.4 and item 4.5 of the “Guidelines for quality criteria in the National Statistical System”*)

For each of the quality components, the following comments could be made:

- The approaches for measuring “Relevance” are one and the same for all types of surveys;
- “Timeliness” for most of the indicators is defined in the relevant legal documents. For surveys based on administrative sources, statistical compilations and combination of data from different sources, “Timeliness” depends on the timely access to the relevant administrative data source;
- There is no significant difference in the approaches used for measuring “Accessibility and Clarity” for the different types of surveys. Important is the information in advance as regards revisions of the main indicators, as well as for assessing differences occurred in the revised indicators. Confidentiality should also be taken into account in announcing data from surveys on enterprises or on individuals carried out on the field, both for exhaustive and sample surveys;
- “Coherence and Comparability” of time series is considered as a main statistical product. Maintaining comparable time series is very important for the users. Interruption of time series requires detailed description of the reasons due and the possibilities this statistical information to be used appropriately. Of major importance is maintenance of the so-called classical time series on specific basic indicators in absolute or relative quantity. As regards geographic comparability, the focus should be on the international data comparison with priority of that within the European Union. Main reasons for causing interruption of comparability at regional level are related to the survey design and the sample design in particular. Comparability between different domains depends on the classifications used at different stages. Of major importance is the coherence of data on separate economic sectors and their incorporation in the system of national accounts, as well as usage of unified definitions, classifications and standards in provision of information for surveys based on combination of data from different sources.

5 CONCLUSION

In applying new integrated approaches for the production of European statistical information, the adoption of the process approach is considered as the most appropriate one. By using the process

approach, efficiency will be increased, burden will be reduced and quality of the statistical information will be improved.

Assessment and monitoring of quality of the statistical process by phases and by sub-processes in different types of surveys will create conditions for improving the management of the statistical process quality and data compilation quality in the future.

Implementation of the approved European Standard on Quality Reports (ESQR) will bring to a successful application of a unified approach in the provision of comparable information on quality in accordance with the European requirements.

6 ANNEXES

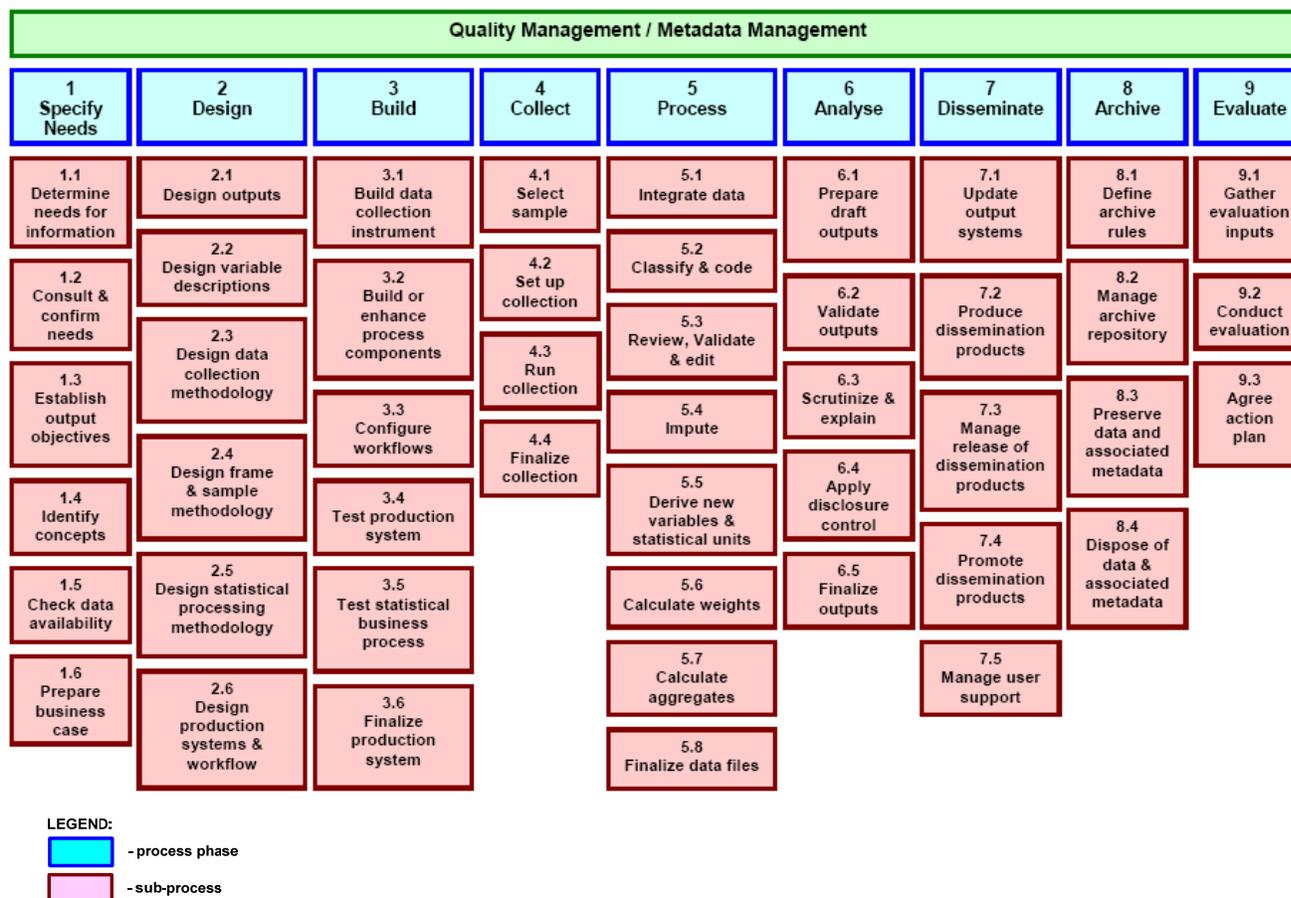
Annex 1: Generic Statistical Business Process Model (GSBPM)

Annex 2: List of process quality indicators and documents

Annex 3: Structure and recommendations for preparing Quality Report

Annex 1: Generic Statistical Business Process Model (GSBPM)

GENERIC STATISTICAL BUSINESS PROCESS MODEL (GSBPM)⁷



⁷ This current version of the model (version 4.0), was approved by the METIS Steering Group for public release in April 2009

Annex 2: List of process quality indicators and documents⁸

Process phases and sub-processes	Indicators / Documents
1. Specify Needs	
1.1. Determine needs for information	1. Contacts with national and international users (number) 2. User requests for existing and new data (number) 3. Users who showed interest in different statistical domains (number) 4. User proposals for inclusion/dropping out of statistical surveys (number) 5. <i>Key users list/database</i> 6. Survey renewal (% of new surveys)
1.2. Consult & confirm needs	1. User consultations carried out (number) 2. Proposals included for new surveys or changes (number)
1.3. Establish output objectives	1. Proposals made for changes in surveys regarding conceptions, time and ways of presentation of data and other characteristics (number) 2. <i>Methodological documents containing the purposes of surveys</i>
1.4. Identify concepts	1. Proposals accepted for changes in surveys regarding conceptions, time and ways of presentation of data and other characteristics (number)
1.5. Check data availability	1. Performed/non-performed data requests, by reasons (number) 2. <i>List of data available in administrative sources which could be used as a resource in realization of the non-performed requests</i> 3. <i>Reference to administrative source data provided under agreements (by kind of data, frequency, source)</i>
1.6. Prepare business case	1. <i>List of tasks</i> 2. Proposals for changes of tasks (number)
2. Design	
2.1. Design outputs	1. <i>List of products (final and intermediate)</i>

⁸ List of process quality indicators is an open document and subject to update

Process phases and sub-processes	Indicators / Documents
	<i>2. List of dissemination methods</i>
2.2. Design variable description	<i>1. List of indicators, incl. metadata</i>
	<i>2. List of standard statistical indicators disseminated free of charge</i>
2.3. Design data collection methodology	1. Studied legal documents (number)
	2. Studied/applied good practices (number)
	3. Used standard concepts, definitions and classifications
	4. Used administrative sources
	5. Data exchange agreements
2.4. Design frame & sample methodology	<i>1. Model and coverage of the sample</i>
	2. Correlation between sample coverage and expenditure on field work and accuracy
	3. Usage of rotation schemes to reduce the response burden
	4. Expenses per unit included in effective sample size (BGN)
	5. Time gap between the reference period to the last update of the sampling frame (months/years)
	<i>6. Usage of software to develop the samples, by type</i>
	7. Difference between the target and achieved sample size
	8. Is substitution allowed in the case of missing units?
	9. Appraisalment of the efficiency of the survey design
2.5. Design statistical processing methodology	<i>1. Technical specification</i>
	<i>2. Schedule</i>
2.6. Design production system & workflow	<i>1. Technical specification</i>
	<i>2. Schedule</i>
3. Build	
3.1 Build data collection instrument	<i>1. Data collection techniques by kind</i>
	<i>2. Usage of administrative data or registers</i>
	<i>3. Approaches used in questionnaire design</i>
	4. Appraisalment of response burden considering the length for responding or completion (minutes/hours)
	5. Appraisalment of response burden considering the compilation of questionnaire or interview
	6. Appraisalment of the quality of the questionnaire design
	<i>7. Description of the results from questionnaire testing</i>

Process phases and sub-processes	Indicators / Documents
3.2 Build or enhance process components	1. Newly developed or updated software applications (number)
3.3 Configure workflows	<i>1. Schedules</i>
3.4 Test production system	1. Undertaken pilot testing (number) <i>2. Report on test results</i>
3.5 Test statistical business process	1. Undertaken pilot testing (number) <i>2. Report on test results</i>
3.6 Finalize production system	<i>1. Documents, service manuals</i> 2. Trainings carried out (number)
4. Collect	
4.1 Select sample	<i>1. List of observed units in the sample</i>
4.2 Set up collection	<i>1. Data collection schedule</i> 2. Punctuality of time schedule (number of delays) 3. Used administrative data and registers 4. Interviewers deployment and training (number) 5. Means of control for the interviewers <i>6. The interview is announced in advance to respondents (by mail, e-mail, etc.)</i> 7. Interviews conducted by one interviewer (average number) 8. Appraisal of training of/by interviewers
4.3. Run collection	1. Timeliness of sending the forms, lists of respondents, notification letters and survey methodologies <i>2. Data collection techniques</i> 3. Effective distribution of the fieldwork 4. Distances that interviewers have to travel to respondents (regarding time and expenses) 5. Time duration of interview/questionnaire filling (minutes/hours) 6. Contacts with respondents (number) 7. Additional contacts aiming at specification of information (number) 8. Correlation between controllers and interviewers on field work
4.4 Finalize collection	<i>1. Report on results</i> 2. Share of responses (%)

Process phases and sub-processes	Indicators / Documents
	3. Appraisalment of overcoverage 4. Appraisalment of undercoverage 5. Response rate of checklist questions 6. Share of refusals (%) 7. Share of respondents who provided information after reminder or additional visit (%) 8. Share of respondents who provided e-information (%) 9. Appraisalment of data collection process 10. Appraisalment of protection of incoming and outgoing microdata
5. Process	
5.1 Integrate data	1. <i>Data input methods</i> 2. Corrections in input data control (number) 3. Corrections in aggregation of primary data (number)
5.2 Classify and code	1. Classification errors 2. Coding errors (number)
5.3 Review, validate and edit	1. Errors made, by kind (number; %) 2. <i>Used data editing procedures</i> 3. Share of corrected errors (%) 4. Outlier detected (number)
5.4 Impute	1. Missing data by reasons (%) 2. <i>Imputation methods used</i> 3. Degree of missing data imputation (%)
5.5 Derive new variables and statistical units	1. Derivative statistical indicators (number) 2. <i>Methods used to form statistical units by aggregation or break down of already collected data, by kind</i>
5.6 Calculate weights	
5.7 Calculate aggregates	
5.8 Finalize data files	
6. Analyse	
6.1 Prepare draft outputs	1. <i>Development of statistical characteristics, major and additional indicators, time series, incl. seasonal adjustment,</i>

Process phases and sub-processes	Indicators / Documents
	<i>indices, etc.</i>
6.2 Validate outputs	1. Comparison with data from past periods and data from other sources 2. Contradictions/deviations 3. Macro editing
6.3 Scrutinize and explain	
6.4 Apply disclosure control	1. Identification of cases of primary and secondary confidentiality and application of procedures for control in announcement 2. Signed agreements for transmission of individual data for scientific purposes (number)
6.5 Finalize outputs	1. Finalize the control on coordination of the results 2. <i>Development of internal documentation</i> 3. Appraisalment of product quality, by components 4. <i>Quality reports</i> 5. Meetings with domain experts and employees dealing with processing and analysis of the results (number) 6. <i>Press releases</i>
7. Disseminate	
7.1 Update output systems	1. Updates (number, %)
7.2 Produce dissemination products	1. Information published on Internet
	2. The NSI website visits (number)
	3. Allocation of the NSI website visits by themes and groups of users (%)
	4. Web-pages in the NSI website (number)
	5. New or updated web-pages (number)
	6. New statistical sections on the NSI website (number)
	7. Time series released on the NSI website (number)
	8. Visits at databases and structures by domains and groups of users (number)
	9. Press releases (number)
	10. Press conferences (number)
	11. Publications (number)
	12. Comments published in media on the results from the NSI statistical surveys (number)

Process phases and sub-processes	Indicators / Documents
	13. Library – number of users served and number of library units
7.3 Manage release of dissemination products	<i>1. Rules for dissemination of statistical products and services</i> <i>2. Release calendar</i> 3. Agreements on data dissemination (number)
7.4 Promote dissemination products	1. Press conferences (number) 2. Issued leaflets, pamphlets, specialized subject materials (number)
7.5 Manage user support	1. Requests for data provision (number) 2. Distribution of the requests for data provision by themes and groups of users – % 3. ESDS Help desk – number of users served 4. ESDS Help desk - Distribution of the requests by themes and groups of users - % 5. Requests via telephone (number) 6. Answers to the requests and consultations via telephone (%)
8. Archive	
8.1 Define archive rules	<i>1. Archiving policy and rules</i> <i>2. Technical and organizational means</i>
8.2 Manage archive repository	<i>1. Catalogue of the archived data and metadata</i> <i>2. Control on uniformity of the archived data and metadata</i>
8.3 Preserve data and associated metadata	1. Archived data and metadata (number)
8.4 Dispose of data and associated metadata	1. Inventory of the archived data and metadata (number)
9. Evaluate	
9.1. Gather evaluation inputs	1. Appraisal and taking account of users opinion 2. Appraisal and taking account of respondent opinion 3. Appraisal and taking account of personnel opinion
9.2. Conduct evaluation	<i>1. Summary reports and proposals for improvement of the process</i> <i>2. Quality management report based on assessments of the nine phases of the process</i> <i>3. Quality management report: internal/external audits on quality</i>

Process phases and sub-processes	Indicators / Documents
9.3. Agree action plan	<i>1. Recommendations for improvement of the statistical process/product quality</i>
	<i>2. Monitoring of the realized recommendations</i>

Annex 3: Structure and recommendations for preparing Quality Report according to the ESS Standard for Quality Reports, 2009

STRUCTURE AND RECOMMENDATIONS FOR PREPARING QUALITY REPORT⁹

1. INTRODUCTION

An introduction is needed to provide context for the quality report.

What should be included in the Introduction to the Statistical Process and Its Outputs?

- A brief history of the statistical process and outputs in question.
- The broad statistical domain to which the outputs belong; related statistical outputs.
- The boundary of the quality report at hand and references to related quality reports.
- An overview of all output produced by the statistical process.
- Reference to other documentation, especially on methodology.

2. RELEVANCE

What should be included on Relevance?

- A content-oriented description of all statistical outputs.
- Definitions of statistical target concepts (population, definition of units and aggregation formula) including discrepancies from ESS/international concepts. (Can also be discussed under Coherence and Comparability.)
- Information on completeness compared with relevant regulations/guidelines.
- Unmet user needs, including reasons for not meeting them.
- Available quality indicators.

3. ACCURACY

3.1. For All Types of Statistical Process

What should be included on Overall Accuracy?

- A presentation of the methodology sufficient for (i) judging whether it lives up to internationally accepted standards and best practice and (ii) enabling the reader to understand specific error assessments.
- Identification of the main sources of error for the main variables.
- A summary assessment of all sources of error with special focus on the key estimates.
- An assessment of the potential for bias (sign and order of magnitude) for each key indicator in quantitative or qualitative terms.

⁹ For further information see ESS Standard and Handbook for Quality Reports, 2009:

http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf

http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQOR_FINAL.pdf

3.2. For Sample Surveys

3.2.1 Sampling errors

What should be included on Sampling Errors?

Always applicable for sampling errors

- Where sampling is used there should always be a section on sampling errors.
- As far as possible sampling error should be presented also for estimates of change in addition to estimates of level. If necessary, reasonable assumptions can be used.

If probability sampling is used:

- Presentation of sampling errors, calculated according to formulas, which should also be made available.
- The most appropriate presentational device should be chosen, normally CV, ranges of CV or confidence intervals.
- If outliers have received special treatment in estimation, this must be clearly described.

If non-probability sampling is used:

- For cut-off sampling an assessment of sampling bias should be included in addition to the presentation of sampling error for the sampled portion of the population.
- For other forms of non-probability a sampling model could be invoked for the estimation of sampling error. A motivation for the chosen model and a discussion of sampling bias should be included.

3.2.2 Coverage Errors

What should be included on Coverage Errors?

- Quantitative information on overcoverage and multiple listing
- An assessment, preferably quantitative, on the extent of undercoverage and the bias risks associated with it.
- Actions taken for reduction of undercoverage and associated bias risks,
- Information on the frame: reference period, updating actions, references to other documents on frame quality.

3.2.3 Measurement Errors

What should be included on Measurement Errors?

- Identification and general assessment of the main risks in terms of measurement error.
- If available, assessments based on comparisons with external data, re-interviews or experiments.
- Information on failure rates during data editing.
- The efforts made in questionnaire design and testing, information on interviewer training and other work on error reduction.
- Questionnaires used should be annexed (if very long by hyperlink)

3.2.4 Nonresponse Errors

What should be included on Nonresponse Errors?

- Nonresponse rates according to the most relevant definitions for the whole survey and for important sub-domains
- Item nonresponse rates for key variables
- A breakdown of nonrespondents according to cause for nonresponse
- A qualitative statement on the bias risks associated with nonresponse
- Measures to reduce nonresponse
- Technical treatment of nonresponse at the estimation stage.

3.2.5 Processing Errors

What should be included on Processing Errors?

- Identification of the main issues regarding processing errors for the statistical process and its outputs.
- Where relevant and available, an analysis of processing errors affecting individual observations should be presented, else a qualitative assessment should be included.

3.3. For Censuses

What should be included on Accuracy for a Census?

- An evaluation/assessment of undercoverage and overcoverage.
- A description of methods used to correct for undercoverage and overcoverage.
- An evaluation/assessment of measurement and classification errors.
- An evaluation/assessment of processing errors, especially where manual coding of data in free text format is used.

3.4. Statistical Processes Using Administrative Source(s)

What should be included on Accuracy for a Statistical Process Using Administrative Source(s)?

- An evaluation/assessment of undercoverage and overcoverage.
- An evaluation/assessment of errors in classification variables.
- For event-reporting systems, an estimate/assessment of the rate of unreported events.

3.5. For Statistical Processes Involving Multiple Data Sources

What should be included on Accuracy for a Statistical Process Involving Multiple Data Sources?

- An overall description of the organisation of the survey, the various segments and a summary of the quality aspects.
- For each segment, the items as specified in the appropriate sections in the Standard. (These items may be grouped by segment and/or by error type/quality component.)

3.6. For Price and Other Economic Index Processes

What should be included on Accuracy for Price or Other Economic Index Process?

- Information on all sampling dimensions (for weights, products, outlets/companies, etc).
- Any attempt at estimating or assessing the sampling error in all or some of these dimensions.
- Quality adjustment methods (including replacement and re-sampling rules) for at least major product groups.
- Assessment of other types of error, where they could have a significant influence.

3.7. For Statistical Compilations

What should be included on Accuracy for a Statistical Compilation?

- Information and indicators relating to accuracy, for example as defined in the IMF's Data Quality Assessment Framework (DQAF) or other relevant, well accepted standard.
- Analysis of revisions between successively published estimates.

For National Accounts

- Analysis of the causes for the statistical discrepancy.
- Assessment of non-observed economy.

3.8. Some Special Issues Concerning Accuracy

3.8.1 Model Assumptions and Associated Errors

What should be included on Model Assumptions and Associated Errors?

- Models related to a specific source of error should be presented in the section concerned.
- Domain specific models, for example, as needed to define the target of estimation itself, should be thoroughly described and their validity for the data at hand assessed.

3.8.2 Seasonal adjustment

What should be included in the Quality Report on Seasonal Adjustment?

- A short description of the method used.
- A report on quality aspects in line with the ESS guidelines on seasonal adjustment.

3.8.3 Imputation

What should be included in the Quality Report on Imputation?

- Information on the extent to which imputation is used.
- A short description of the methods used and their effects on the estimates.
(Typically this information will be reported in the section(s) dealing with the errors that imputation is helping to correct rather than in a separate section.)

3.8.4 Mistakes

What should be included on Mistakes?

- The number and nature of mistakes over the past few years should be described.
- Measures taken to avoid mistakes in the future should be described.

3.8.5 Revisions

What should be included on Revisions?

- The revision policy.
- The average number of revisions (planned and unplanned).
- The average size of revisions.
- The main reasons for revisions, and the extent to which the revisions improved accuracy of statistics.

4. TIMELINESS AND PUNCTUALITY

What should be included on Timeliness and Punctuality?

- For annual or more frequent releases: the average production time for each release of data.
- For annual or more frequent releases: the percentage of releases delivered on time, based on scheduled release dates.
- The reasons for non-punctual releases explained

5. ACCESSIBILITY AND CLARITY

What should be included on Accessibility and Clarity?

- A description of the conditions of access to data: media, support, pricing policies, possible restrictions etc.
- A summary description of the information (metadata) accompanying the statistics (documentation, explanation, quality limitations, etc).
- The description should refer to both less sophisticated and more advanced users and how their needs have been taken into account.
- A summary of user feedback on accessibility and clarity.

6. COHERENCE AND COMPARABILITY

What should be included on Coherence and Comparability?

General:

- Brief descriptions of all conceptual and methodological metadata elements that could affect coherence/ comparability.
- An assessment (preferably quantitative) of the possible effect of each reported difference on the output values.

- Differences between the statistical process and the corresponding European regulations/ standards and/or international standards (if any).
- Comparability over Time:
- Reference periods when series breaks (if any) occurred, the reasons for them and treatments of them.
- Comparability over Region:
- A quantitative assessment of comparability across regions based on the (weighted) number of differences in metadata elements.
 - At ESS level, a coherence/comparability matrix summarising the possible sources of lack of comparability with any specified standard by region.
- Internal Coherence
- Any lack of coherence in the output of the statistical process itself.
- Coherence with National Accounts
- Where relevant, the results of comparisons with National Account framework and feedback from National Accounts with respect to coherence and accuracy problems.
- Coherence with Other Statistics
- Where the statistical outputs were combined with those from other processes, the limitations if any set by coherence considerations
- Mirror Statistics
- Assessment of discrepancies (if any).

7. TRADE-OFFS BETWEEN OUTPUT QUALITY COMPONENTS

What should be included on Trade-offs?

- A description of each important trade-off that has been analysed and the basis on which the trade-off decision has been made.
- A statement concerning any trade-offs that should have been analysed but have not been.

8. ASSESSMENT OF USER NEEDS AND PERCEPTIONS

What should be included on User Needs and Perceptions?

- Means of obtaining information on users and uses.
- Description and classification of users.
- Uses for which users want the outputs.
- Users and uses given special consideration.
- Means of obtaining user views.
- Main results regarding user satisfaction.
- Date of most recent user satisfaction survey.

9. PERFORMANCE, COST AND RESPONDENT BURDEN

What should be included on Performance, Cost and Respondent Burden?

Performance and Cost:

- Annual operational cost with breakdown by major cost component.
- Recent efforts made to improve efficiency.
- The procedures for internal assessment and for independent external assessment of efficiency.
- The extent to which routine operations, in particular data capture, coding, validation and imputation, are automated.
- The extent to which ICT is effectively used for data collection and dissemination and the improvements that could be made.

Respondent Burden

- Annual respondent burden in financial terms and/or hours.
- Respondent burden reduction targets.
- Recent efforts made to reduce respondent burden.
- Whether the range and detail of data collected by survey is limited to what is absolutely necessary.
- Whether administrative and other survey sources are used to the fullest extent possible.
- The extent to which data sought from businesses is readily available from their accounts.
- Whether electronic means are used to facilitate data collection.
- Whether best estimates and approximations are accepted when exact details are not readily available.
- Whether reporting burden on individual respondents is limited to the extent possible by minimizing the overlap with other surveys.

10. CONFIDENTIALITY, TRANSPARENCY AND SECURITY

What should be included on Confidentiality, Transparency and Security?

Confidentiality:

- Whether or not confidentiality is required by law and if so whether survey staff have signed legal confidentiality commitments.
- Whether external users may access micro-data for research purposes, and, if so, the confidentiality provisions that are applied.
- The procedures for ensuring confidentiality during collection, processing and dissemination, including rules for determining confidential cells in output tables and procedures for detecting and preventing residual disclosure.

Transparency:

- The ways in which the uses to which the data are to be put and confidentiality provisions are made known to respondents.

- Whether statistical announcements and statements made in press conferences are objective and non-partisan.
- Whether errors discovered in published statistics are corrected and publicised.

Security

- The provisions in place to ensure the security of data acquisition processes, in particular data collected through the Internet.
- The provisions that are in place to ensure the security and integrity of completed questionnaires, micro and macro databases and data outputs.

11. CONCLUSION

What should be included in the Conclusion?

- Principal quality problems.
- Recommendations for improvements.
- Follow-up action items.

7 REFERENCES

1. Burg T. (2010), “Are the Standard Documentation really Quality Reports?”, presented at the Q2010 Conference, Helsinki, May 2010.
2. Eurostat (2007), Handbook on Data Quality Assessment Methods and Tools (DatQAM).
3. Eurostat (2008), Handbook on improving quality by analysis of process variables.
4. Eurostat (2009), ESS Standard for Quality Reports.
5. Eurostat (2009), ESS Handbook for Quality Reports.
6. National Statistical Institute of Bulgaria (2005), “Compendium of methodology for quality assessment of statistical data in Bulgaria”, prepared in the framework of a National Phare Project 2002.
7. National Statistical Institute of Bulgaria (2005), „Instructions for implementation of the quality reports of statistical surveys”, prepared in the framework of a National Phare Project 2002.
8. National Statistical Institute of Bulgaria (2008), “Process quality indicators”, prepared in the framework of Grant Project “Definition of process variables and development of standards”.
9. UNECE (2009), Generic Statistical Business Process Model, Version 4.0, prepared by UNECE Secretariat and approved by the METIS Steering Group for public release in April 2009.