REPUBLIC OF BULGARIA
National Statistical Institute

"ENVIRONMENTAL GOODS AND SERVICES" STATISTICAL SURVEY METHODOLOGY

The statistical survey is conducted annually by the Environmental and Energy Accounts Department, Macroeconomic Statistics Directorate, of the National Statistical Institute, on the basis of mandatory participation, according to the National Statistical Program.
According to Art. 20 of the Statistics Act, respondents are obliged to provide the National Statistical Institute with reliable data on surveys included in the National Statistical Program, which are intended to be conducted on the basis of mandatory participation.
According to Art. 25, para. 1 of the Statistics Act, individual data obtained and collected during statistical surveys are confidential and can only be used for statistical purposes.

GENERAL INFORMATION

OBJECTIVE
The main purpose of the statistical survey is to provide information to all stakeholders and the public about the produced environmental goods, services and technologies related to environmental protection and natural resource management, in accordance with Annex V of Regulation (EU) № 691/2011 on the European Environmental Economic Accounts for the module for the Environmental goods and services sector accounts.
The main purpose of the environmental goods and services sector accounts is to record and present data on the production operations of national economies resulting in products and services related to environmental protection and natural resource management in a manner consistent with the reported data, according to the European System of Accounts (ESA) 2010. These accounts should use already existing information from national accounts, structural business statistics, foreign trade and other data sources.

ESSENCE
The environmental goods and services sector (EGSS) occupies an important place in the analyzes of the green economy and policies for the environment and natural resources and for monitoring the Sustainable Development Goals (SDGs).
The environmental goods and services sector has the same coverage, i.e. the same system boundaries as the ESA and consists of all environmental goods and services that are created within production. ESA defines production as the activity carried out under the control and responsibility of an institutional unit that uses input of labour, capital, goods and services to produce output of goods and services.
Environmental products are the products produced with main purpose of environmental protection and resource managing. Resource management involves the conservation, maintenance and improvement of the stock of natural resources, and therefore the protection of these resources against depletion.
In order to be defined as such, environmental technologies, products and services must satisfy the main (environmental) purpose criterion, which is in this case environmental protection or resource management. This is mainly determined on the basis of the technical nature of the activity or the intention of the manufacturer.

Environmental goods and services sector accounts are in accordance with the accounting structures and principles of the System of Environmental Economic Accounting - Central Framework - SEEA CF. SEEA CF is an international standard for concepts, definitions, classifications, accounting rules and tables for the preparation of internationally comparable statistics on the environment and its relationship to the economy.

PURPOSE

The collected data from Environmental goods and services statistical survey are used for the annual reporting to the EC - DG Eurostat by filling in an electronic excel questionnaire for Environmental goods and services sector accounts, according to Annex 5 to Regulation (EC) No. 691/2011 (amended by Delegate Commission Regulation (EU) 2022/125).

The data are published annually, according to the Release Calendar, on the NSI website for public information.

STATISTICAL SURVEY

SCOPE, STATISTICAL UNIT AND GENERAL POPULATION

The statistical survey covers enterprises and kind-of-activity units (KAUs) from all economic activities.

From the "General Government" sector (S13), according to the ESA 2010 (paragraph 2.113) the institutional units classified in the following sub-sectors are covered:

- Central government - S.1311 (includes all units which accounts are included in the central budget; Higher education institutions; Bulgarian Academy of Sciences; Bulgarian National Television; Bulgarian National Radio; extra-budgetary accounts of central government; hospitals financed by the central government and other institutional units with independent budgets);
- Local government - S.1313 (includes all municipalities; the hospitals financed by them and extra-budgetary accounts and funds to the municipalities);
- Social security funds- S.1314 (includes the National Insurance Institute and the National Health Insurance Fund).

The statistical unit is an enterprise and kind-of-activity unit. For the sector "General Government" is an institutional unit.

The statistical population consists of active enterprises, kind-of-activity units from all economic activities and institutional units from the "General Government" sector.

The statistical survey is a combination of comprehensive and sample survey.

DATA SOURCES

- Statistical survey ‘Environmental goods and services’.

Information from other statistical surveys (business statistics, non-financial national accounts, foreign trade, agriculture and forestry, energy statistics) and administrative sources (MAF, SEDA) is also used to define the population, verify the data and make estimates.

FREQUENCY OF DATA COLLECTION AND PUBLICATION

Annually.

STATISTICAL STUDY

DEFINING THE STATISTICAL POPULATION FOR THE STUDY

The statistical population is determined by statistical units (enterprises, kind-of-activity units from all economic activities and institutional units from the "Central Government" sector with more than 10 employed persons included in the Statistical Business Register (SBR). Statistical units with less than 10 employees are excluded from the surveys, in order to reduce the respondents' burden. Units with less than 10 employees are estimated.

The selection of the statistical units from the SBR takes into account the type of the survey, which includes a combination of comprehensive and sample statistical survey.

The list of statistical units included in the statistical population contains:

- Statistical units for comprehensive survey. Administrative sources and registers are used to determine them. Additionally, units for which we have information in advance from previous statistical surveys that they are producers of environmental goods and services, are also included.
- A simple random sample selection for the part with the non-covered statistical units.

Selection of simple random samples for the part of non-covered statistical units, due to the large number of enterprises. A separate sample is selected for each division by NACE Rev.2. As the reporting requirements impose disaggregation of the populations, it turns out that in some divisions the number of surveyed units is too large for comprehensive survey but small for sampling. In this case, units are surveyed comprehensively.

To calculate the sample size, the following is determined:

- The guarantee probability to guarantee the survey results and the related guarantee multiplier;
- The maximum size of the stochastic error;
- The standard deviation of employed persons.

A 95% confidence interval is chosen. Practically, this means that the confidence interval will contain the real population average with 95% probability.

The sample size is calculated using the formula:

$$n=\frac{t^2.(V\%)^2}{(\Delta\%)^2+\frac{t^2.(V\%)^2}{N}}$$

where:

Δ%	-	the maximum error;
t	-	guarantee multiplier (1.96 for an interval in which the actual value of the mean is found with a 95% probability, and 2.58 for a 99% probability, respectively)
v%	-	coefficient of variation in % (V%=100*standard deviation / average value - the arithmetic mean)
n	-	number of cases in the sample;
N	-	number of cases in the study population.

After determining the statistical population, compiling the list of all population units and calculating the sample size, simple random samples in each division by NACE Rev.2 are selected, using the SPSS software product.

The two lists are merged - the list of statistical units for comprehensive survey and the list of units selected by simple random sampling.

The population is updated annually before the start of the Campaign.

DATA COLLECTION

Primary data are collected from respondents for the previous reporting year. The data are collected via Environmental Statistics Information System (ESIS), which includes automatic checks for completeness, valid values, and logical data control, according to the specifics of the survey. The nomenclatures used for its needs and instructions for the respondents are available in the system.

PRIMARY DATA PROCESSING, DATABASE PREPARATION

PRIMARY DATA PROCESSING

After the respondents have reported their data in ESIS, their processing begins. It is carried out in the system according to the survey program schedule and goes through several stages.

Data validation is carried out on the basis of clearly defined criteria regarding their completeness, correct classification, units of measure, comparability with previous years, logical control, etc.

All checks are done at the respondent level.

DATABASE PREPARATION

After finalising the process of removing all identified errors and discrepancies in the data, actions on classifying and coding the data related to the statistical units or collected variables are also taken. The purpose of these encoding procedures is to obtain derived variables, numerical values, or aggregate values during the next processing stages. The coding process is carried out using code tables. When new versions of the used classifications and nomenclatures appear, recoding and reclassification of statistical units and variables is carried out.

After completing these actions, the methodologists prepare the primary database for further processing and calculation the statistical data.

CALCULATION AND ANALYSIS OF STATISTICAL INDICATORS

DATA PREPARATION

Data preparation is a set of processes through which the primary data are brought into a form suitable for the statistical data calculation, analysis and assessment of their quality and the subsequent presentation to users, incl. fulfillment of reporting obligations to Eurostat, according to the specified table formats.

ASSESSMENT OF PRIMARY DATA RESPONSE RATE

Reports from all survey respondents should be collected within the Campaign. In case a 100% response rate is not achieved, the need for an estimates of the missing data is assessed and, if necessary, carried out.

STATISTICAL DATA CALCULATION

From the database with final primary data, the following indicators are calculated: produced output (market output), added value, non-market output of the units from the ‘General Government’ sector.

The market output of environmental goods and services is calculated as the difference of net revenues of sales minus balance value of sold goods.

The value added of market output of environmental goods and services is the difference of market output minus intermediate consumption.

Intermediate consumption is calculated as the sum of current costs: expenditures for raw materials plus expenditures for external services plus other expenditures (excluding balance value of sold assets and provisions) plus expenditures for business trips.
The non-market output of environmental services for the units from the 'General Government' Sector is the sum of all current operational costs (maintenance: including expenditures of raw materials and external services plus labor and social security costs).
All parameters of the population are estimated based on the sample data.
In the divisions by NACE Rev.2 in which a sample is selected, an estimate is calculated by multiplying the average number of the relevant characteristic of an enterprise in the sample by the number of enterprises in the division with more than 10 employees from the population. Thus, the estimated total value of the relevant parameters, for which the population is surveyed by sample, is obtained. The estimation is done only for the sample data. The sample estimate is burdened with error, i.e. the estimated total value of the indicators is in the middle of the confidence interval, which is calculated as follows: the estimated total value of the corresponding variable from the sample data -/+ the maximum error (the stochastic error multiplied by the guarantee factor).
The maximum error is calculated using the following formula:

$$\Delta\% = \sqrt{\frac{t^2 V\%^2}{n} - \frac{t^2 V\%^2}{N}}$$

where:

Δ%	the maximum error;
t	guarantee multiplier (1.96 for an interval in which the actual value of the mean lies with a probability of 95%,)
v%	coefficient of variation in % (V%=100*sample standard deviation/sample mean, respectively Stdev and Average in Excel)
n	number of cases in the sample;
N	number of cases in the study population.

To obtain the data of all enterprises in the population, the data estimate of the sampled population is added to the data of the enterprises of the comprehensively studied population, where it is available for the relevant division.

In order to also estimate the data of enterprises with less than 10 employees, the share of employees in enterprises with 10 or more employees from those employed in all enterprises from the population in the relevant division is calculated (formula 1).

Formula 1:

$$D_{\geq 10} = \frac{\sum V16110_{\geq 10}}{\sum V16110} * 100 ,$$

where:

$D_{\geq 10}$	-	share of employees in enterprises with 10 or more employees of those employed in all enterprises of the general population in the relevant industry group;
$\sum V16110_{\geq 10}$	-	number of employees in enterprises with 10 or more employees in the relevant industry group;
$\sum V16110$	-	number of employees in all enterprises from the general population in the relevant industry group.

The data of the relevant indicators in enterprises with less than 10 employees is calculated according to formula 2.

Formula 2:

$$P_{\leq 9} = \frac{\sum P_{\geq 10}}{D_{\geq 10}} * (100 - D_{\geq 10})$$

where:

$P_{<9}$ - assessment of the indicators of enterprises with less than 10 employees;

$\sum P_{\geq 10}$ - data of enterprises with 10 or more employees;

$D_{\geq 10}$ - share of employees in enterprises with 10 or more employees of those employed in all enterprises from the general population in the relevant industry group.

The estimation of the relevant indicators in the enterprises with less than 10 employees is added to the estimated total value of the indicators in the enterprises with more than 10 employees, for which the population is examined by simple random sampling.

CONFIDENTIALITY

According to the Statistics Act, Art. 25, para. 1, item 3: 'The National Statistical Institute and statistical authorities and their employees may not publicize or provide statistical information that summarizes data for less than three statistical units or for a population in which the relative share of the value of a studied parameter of one unit is more than 85 percent of the total value of this parameter for all units of the population.' For this reason, two types of restrictions apply to the processing and summarization of the data from the 'Environmental goods and services' survey:

- Threshold rule;
- Dominant (from English dominance) rule or also called (n,k) rule.

The threshold rule ensures that if the value of a given cell from the statistics table is formed by 1 or 2 statistical units that cell remains hidden.

In addition to the threshold rule (also known as confidentiality type "A"), the second rule is added - the dominant rule (also known as confidentiality type "B"). After the breakdowns are checked for the presence of at least 3 units in each cell, i.e., the primary confidentiality type "A" is ensured, a verification of confidentiality type "B" is carried out. With the latter, according to the Statistics Act, the rule (1;85) is applied, i.e. the information in a given cell is subject to "suppression" if one statistical unit (respondent) forms 85% or more of the value of the indicator for the whole studied population or subset (eg one enterprise forms 85% of the market output). Survey data tables are hierarchical tables, i.e. the variables that compose them have several values and each value is decomposable into other values. The economic activity of enterprises is a common example of a hierarchical variable that can be decomposed into lower levels. This also requires checking the tables for secondary confidentiality.

Secondary confidentiality is related to the closing of non-confidential values to prevent third parties from indirectly disclosing data from the totals or subtotals for each economic activity by NACE Rev.2 and CEPA/CReMA classes that are determined to be confidential during data verification for primary confidentiality.

The Eurostat software - CIF version 2 is used to calculate the secondary confidentiality.

Specialised software – Excel, SPSS and/or R – is used to calculate the data.

QUALITY ASSESSMENT

The statistical survey follows the General model of the statistical production process in NSI. The quality assessment of statistical data is carried out in order to ensure compliance with the requirements of the Statistics Act. Data quality is ensured by applying the principles of the European Statistics Code of Practice and the NSS Common Framework for Quality Management.

To ensure high quality of the data, their consistency with the data from the database of the annual activity report of enterprises, Foreign trade, Environmental protection and restoration expenditures and Revenues and expenditure for municipal waste and expenditures for water supply infrastructure, as well as with data reported to Eurostat for main macroeconomic indicators output and value added from Non-financial national accounts - Aggregated national accounts by industry (A64 based on NACE, Rev.2).

Efforts are being made to continuously improve the quality of the data, by improving the primary data collection system through the online-based ESIS, ensuring strict arithmetic and logical control of the input data, and by performing additional analyzes and verifications.

With the preparation and reporting of the statistical data, a quality report is also prepared, which is filled in the Eurostat system for metadata and quality reports.

Quality report and metadata are also published on the NSI website together with the statistical data. They are updated annually and contain additional information related to the survey.

STATISTICAL PRESENTATION

The statistical data are reported annually to Eurostat in a standardized Excel format (EGSS Questionnaire) through EDAMIS.

They are published on the NSI website in the 'Statistical data - Environment' section, as well as in the Infostat system. The data are used for the preparation of NSI publications - Statistical Yearbook, Statistical Reference Book, brochure 'Bulgaria', specialised electronic publication 'Environment', as well as for providing information services upon users' request.

The data are presented by groups of economic activities at national level in million BGN.

The Eurostat questionnaire contains data on output (total, market output, non-market output, ancillary output, and own final use), value added, exports and employment.

The data for the environmental goods and services sector are presented by:

- Economic activities and sectors for reporting to EUROSTAT - according to the level of aggregation A21 by NACE Rev. 2 as defined in ESA 2010;
- CEPA and CReMA classes;
- Breakdown by market activities and its characteristics (output, exports, gross value added and employment), as the questionnaire allows the voluntary reporting of two separate categories 'of which' environmental products: environmental specific services and cleaner and resource efficient products for environmental protection or resource management.