BUILDING AN INFORMATION SYSTEM ARCHITECTURE FOR ANALYZING THE CRITERIA OF SUSTAINABLE DEVELOPMENT GOALS

Skvortsova D.A.[1], Filin N.A.[2*], Chupis S.A.[3]

[1]Department of Industrial Logistics,Moscow State Technical University,Moscow,Russian
[2]Department of Industrial Logistics,Moscow State Technical University,Moscow, Russian
[3]Department of Industrial Logistics,Moscow State Technical University,Moscow, Russian

darya_skv@mail.ru[1]
nikitafilin2004@gmail.com[2]
sergei.chupis@mail.ru[3]
*
Corresponding Author:darya_skv@mail.ru

Abstract—It is proposed to measure sustainability by a single super-criterion, the value of which is influenced by sustainability criteria in the social, economic and environmental spheres. The third level includes criteria that determine the solution of 17 tasks for achieving the Sustainable Development Goals, which in turn, at the fourth level, 176 criteria are presented. Among them, a group of statistically significant indicators was identified. The indicators of this group were normalized, and then checked for normality. A correlation matrix of indicators of the Sustainable Development Goals was compiled.The main results are: statistically significant was 38% of the indicators from the entire data set, which were reduced to a single norm by the robust sigmoid function, and correlation coefficients were calculated. At the same time, coefficients with positive values are most fully described by the Gauss law, and distributions of coefficients with negative values are most fully described by the Weibull distribution, which allows us to determine the mutual influence of SDG indicators for the subsequent construction of a data showcase.

A database that is part of the system architecture has been developed. Its implementation allows you to collect user responses for expert assessments when determining weighting coefficients for convolving sustainable development criteria at all graph levels. Also, information about the initial values of indicators, normalized data, and aggregated data will be collected and processed in the corresponding entities, which will allow performing multidimensional analysis in the data showcase on a macro and micro scale using various visualization methods.

Keywords—architecture of an information system for visualizing data on sustainable development, correlation analysis,indicators of the Sustainable Development Goals,sustainable development.

I. INTRODUCTION

The generalized structure of sustainable development criteria should be represented as a four-level cyclic graph, where each node of which corresponds to a certain criterion. It is proposed to measure sustainability by a single supercriterion, the value of which is influenced by sustainability criteria in the social, economic and environmental spheres. The third level includes criteria that determine the solution of 17 tasks for achieving the Sustainable Development Goals, which in turn, at the fourth level, 176 criteria are presented. Based on the graph, a database will be created to form a data showcase for sustainable development analysis in the form of a relational model.

It is necessary to identify a group of statistically significant indicators of the Sustainable Development Goals. Indicators of this group should be normalized, and then checked for normality. To determine the degree of influence of criteria and groups of criteria on each other and on the super-criterion, it is necessary to build a correlation matrix of indicators of the Sustainable Development Goals, in which it is necessary to distinguish two groups depending on the nature of the relationship between the indicators – positive or negative. It is also necessary to calculate the main characteristics – the average value of the correlation coefficient, the lower quartile, the upper quartile, the minimum and maximum values, as well as plot the distribution density.

To build a data showcase, you need to perform data analysis, build a common architecture and database schema of the system for visualizing data on sustainable development.

II. APPLICATION OF CORRELATION ANALYSIS TO DETERMINE THE RELATIONSHIP BETWEEN INDICATORS OF THE SUSTAINABLE DEVELOPMENT GOALS

The generalized structure of sustainable development criteria can be represented as a complex graph consisting of many nodes (vertices) $\{k_i\}$, each of which corresponds to a certain criterion $K_{i,j}$. Connections between nodes are interpreted by its edges.

Sustainable development is defined as the interrelation of many indicators [1]. At the same time, in general, sustainability in this interpretation can be measured by a single super criterion (SI – Super Indicator), the value of which is influenced by sustainability criteria in the social (SIsocial), economic (SIeconomic) and environmental (SIecological) spheres [2]. At the same time, it is a difficult task to determine the degree of influence of each of them on SI at different points in time. There are also several interrelated generalized sets of criteria for sustainableAn easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

There are also several interrelated generalized sets of criteria for sustainable development that relate to different tasks for achieving the goals set. Based on Rosstat data, there are 17 groups of Sustainable Development Goals (SDGs):

1. End poverty in all its forms everywhere.
2. Eliminate hunger, ensure food security and improve nutrition, and promote sustainable agricultural development.
3. Ensure healthy lifestyles and promote well-being for all at all ages.
4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.
5. Promote gender equality and the empowerment of all women and girls.
6. Ensure the availability and rational use of water resources and sanitation for all.
7. Ensuring access to affordable, reliable, sustainable and modern energy sources for all.
8. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.
9. Build resilient infrastructure, promote inclusive and sustainable industrialization and innovation.
10. Reducing inequality within and between countries.
11. Ensuring the openness, security, resilience, and environmental sustainability of cities and localities.
12. Ensuring the transition to rational consumption and production patterns.
13. Take urgent action to combat climate change and its consequences.
14. Conservation and rational use of the oceans, seas and marine resources for sustainable development.
15. Protection and restoration of terrestrial ecosystems and promotion of their rational use, sustainable forest management, combating desertification, stopping and reversing land degradation and stopping the loss of biological diversity.
16. Promote a peaceful and inclusive society for sustainable development, ensure access to justice for all, and build effective, accountable and inclusive institutions at all levels.
17. Strengthening the means of implementation and revitalizing the Global Partnership for Sustainable Development Financing.

(On the one hand, each of these 17 SDG –Sustainable development goals is measurable [3]. On the other hand, the degree of influence of each of them on the value of SIsocial, SIeconomic and SIecological is also unknown. In order to achieve the SDGs, we have formulated targets that are defined by 176 criteria that fall into these 17 groups. It can be used as a designation for 17 SDGs - SDG1, SDG2,..., SDG17. The designations of the criteria for each SDG group are shown in table 1.

TABLE I
DESIGNATIONS OF CRITERIA FOR EACH SDG GROUP

Designation of SDG criteria	**Number of criteria for the group**	**Designations of the SDG criteria group**
*SDG*1	*5*	*SDG*1.1, *SDG*1.2, ..., *SDG*1.5
*SDG*2	*6*	*SDG*2.1, *SDG*2.2, ..., *SDG*2.6
*SDG*3	*39*	*SDG*3.1, *SDG*3.2, ..., *SDG*3.39
*SDG*4	8	*SDG*4.1, *SDG*4.2, ..., *SDG*4.8
*SDG*5	5	*SDG*5.1, *SDG*5.2, ..., *SDG*5.5
*SDG*6	6	*SDG*6.1, *SDG*6.2, ..., *SDG*6.6
*SDG*7	6	*SDG*7.1, *SDG*7.2, ..., *SDG*7.6
*SDG*8	21	*SDG*8.1, *SDG*8.2, ..., *SDG*8.21
*SDG*9	32	*SDG*9.1, *SDG*9.2, ..., *SDG*9.32
*SDG*10	7	*SDG*10.1, *SDG*10.2, ..., *SDG*10.7
*SDG*11	15	*SDG*11.1, *SDG*11.2, ..., *SDG*11.15
*SDG*12	8	*SDG*12.1, *SDG*12.2, ..., *SDG*12.8
*SDG*13	3	*SDG*13.1, *SDG*13.2, *SDG*13.3
*SDG*14	3	*SDG*14.1, *SDG*14.2, *SDG*14.3
*SDG*15	4	*SDG*15.1, *SDG*15.2, *SDG*15.3, *SDG*15.4
*SDG*16	4	*SDG*16.1, *SDG*16.2, *SDG*16.3, *SDG*16.4
*SDG*17	6	*SDG*15.1, *SDG*15.2, *SDG*15.3, *SDG*15.4

Analyzing this set, we can conclude that it is necessary to select the basic elements. They form the root nodes of a generalized graph and, in turn, can form their own sets. In addition, it is advisable to divide components into levels according to their purpose.

Thus, a complex formal graph structure is obtained (Fig. 1), which contains several root nodes, i.e. nodes that do not have parents; edge (hanging) nodes that do not have descendants; cycles in the structure of its parts; branches without cycles. It is advisable to assign unique designations to all elements (nodes of the graph).

According to graph theory, a complex graph with cycles has no unambiguous routes. To make the graph structure unambiguous, you should enter the orientation (direction) of the edges of the corresponding graph, according to the hierarchical subordination of its components. Obviously, the edges can be oriented either from bottom to topor from top to bottom, starting from the root nodes.

The basis is a tree graph that has a single root (parent) node and edges connecting the nodes. In this case, there is a single route between any two nodes in the tree.

However, the sustainable development criteria graph is not a tree graph, but a more complex graph containing formal cycles and branches. The first level criterion is the generalized super-criterion of sustainable development. At the second level, it includes interrelated criteria that determine development in the field of ecology, economy and social sphere. The third level includes criteria that determine the solution of 17 tasks for achieving the Sustainable Development Goals. In turn, at the fourth level, 176 criteria are presented, which are included in 17 groups. Thus, a four-level cyclical graph of sustainable development criteria can be formed, on the basis of which a database will be formed for forming a data showcase for analyzing sustainable development [4].

As can be seen from the graph structure, the data model [5] can be represented as a network model or a relational one. However, when building a network data model that has a clear structure, a large amount of additional information about relationships will be used, which is not optimal when processing a large amount of data. In a relational model, data can be represented as relationships that change over time, and attributes have unique names within a single relationship. It is also convenient to physically implement the relational model.

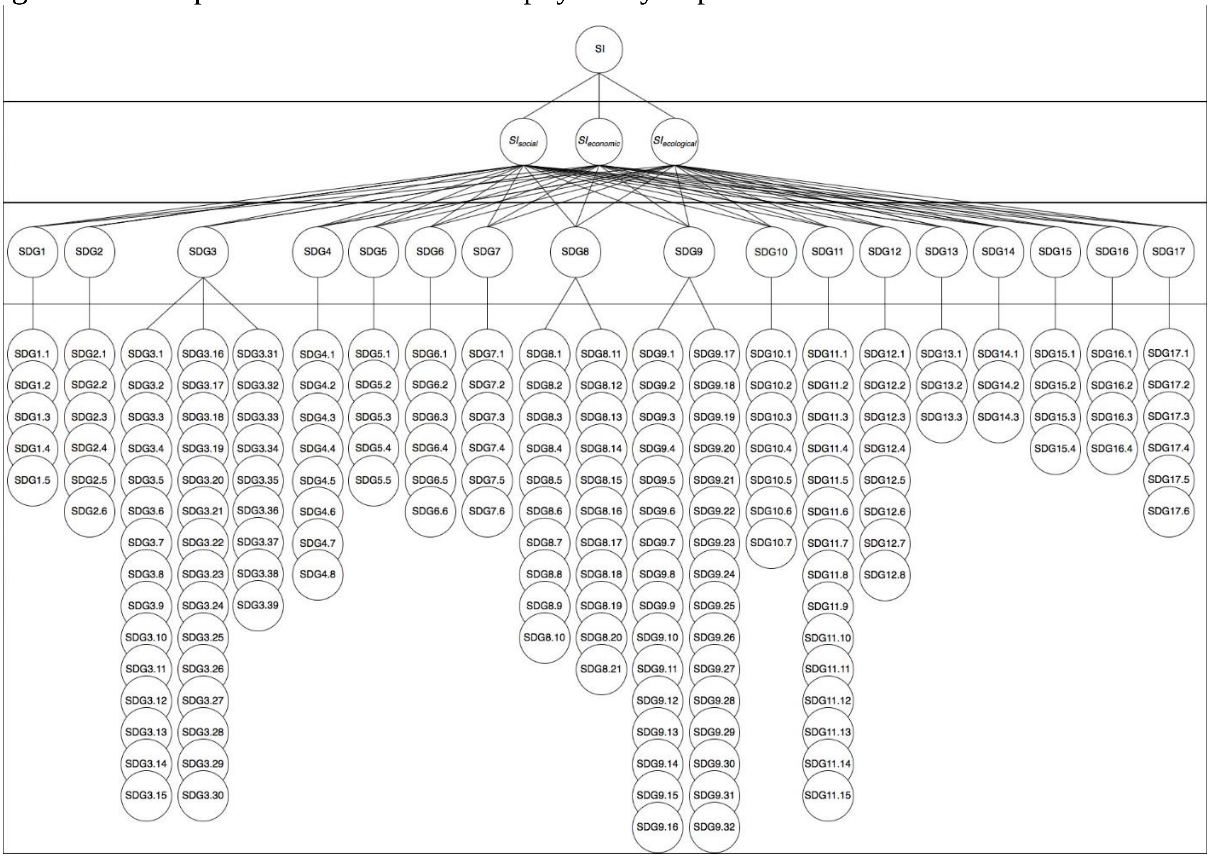

Fig. 1. Formal structure of the SDG graph

One of the options for analyzing and evaluating the effectiveness of indicators of the Sustainable Development Goals can be the linear optimization method: multi-criteria or binary. The use of multi-criteria optimization is acceptable if several groups of indicators are introduced. The use of binary optimization is acceptable when setting a problem with the choice of specific indicators from a set of possible ones. Another group of methods for evaluating the effectiveness of indicators of the Sustainable Development Goals (SDGs) are point-based methods for evaluating effectiveness: additive and multiplicative convolution, and point-based methods for evaluating SDG indicators. The use of these methods can be supplemented with correlation analysis to refine the data.

The r-Pearson correlation coefficient is used to study the relationship between two metric variables measured in the same sample. However, it should only be used if the following conditions are met:

1. both variables are quantitative and continuous,
2. both features have a normal distribution (therefore, calculating this coefficient is a parametric method for evaluating the relationship of features).
3. the relationship between variables is linear,
4. homoscedasticity (the variability of one variable does not depend on the values of another variable),
5. independence of research participants from each other,
6. pair of observations,
7. for an adequate projection of calculations on the general population, the sample should be representative [6].

According to Rosstat, the initial data consists of 176 time series with a periodicity of 1 year and a length of no more than 13 years. Based on this, we can say that with such a sample size with a statistical power of 90% and a 95% confidence interval, we can reject the null hypothesis that the correlation coefficient is zero for various alternative hypotheses, where the value of the correlation coefficient is at least 0.8 [7].

However, time series with a number of dimensions less than 10 significantly limit the statistical power, so they were excluded from the analysis. We also excluded indicators whose measurement methodology has changed, since such indicators are considered insignificant from the econometrics point of view. Changing the calculation methodology makes it difficult to identify long-term trends. This can lead to incorrect conclusions about significance or correlations. Thus, 82 out of 176 indicators were submitted for analysis. Indicator number 129 (Income differentiation coefficient (by 10% population groups)) It was divided into 2 separate indicators, as it contained two separate indicators: 129-1-decile coefficient and 129-2-fund coefficient.

Normalization of the data [8] for subsequent correlation analysis was performed using the robust sigmoid function:

$$\tilde{f} = \left[1 + exp(-\frac{f - m_f}{1.35 r_f})\right], \qquad (1)$$

where $\tilde{f}$ is the normalized feature values in the time series data set, f is the vector of non-normalized feature values, m_f is the median of f and r_f is its interquartile range [9].

Further, the normalized data were checked for normality using the Shapiro-Wilk test [10]. As a result, out of 82 indicators, 67 remained in the data set.

When all the conditions were met, a correlation matrix was compiled for calculating the pairwise r-Pearson correlation coefficients (Fig. 2).

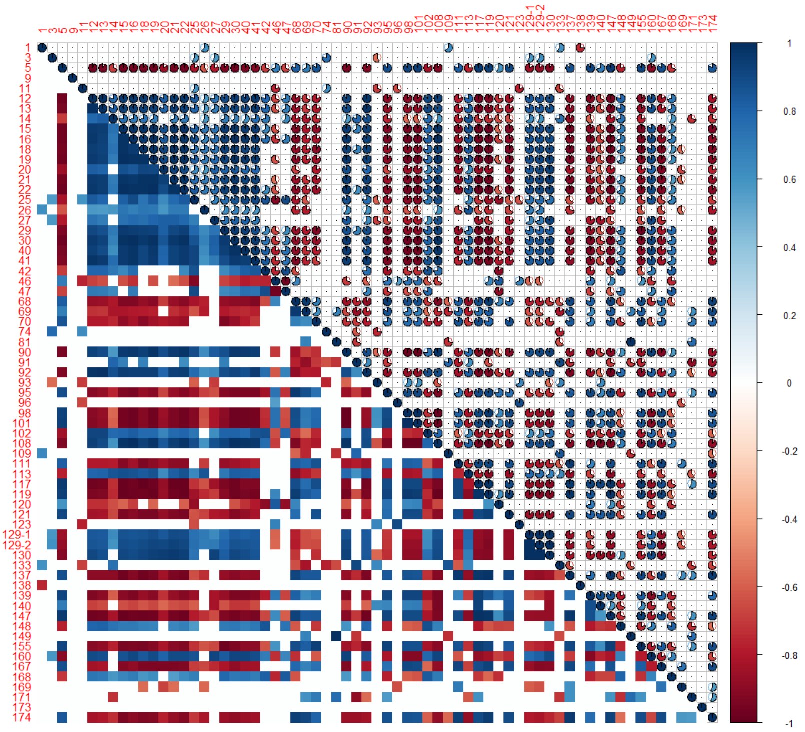

Fig. 2-SDG Correlation matrix

The values of the correlation coefficients in the upper part of the matrix are presented in the form of a pie chart, in which negative values of the correlation coefficients are expressed in red, and positive values are expressed in blue. The lower part of the matrix is visualized as a heat map, where the hue of the color depends on the values of the correlation coefficient, and red is used for the negative value, and blue for the positive value. The minimum negative value is shown in dark red, and the maximum positive value is shown in dark blue.

The resulting set of values of correlation coefficients can be divided into two groups: positive and negative, and for each of these groups, the main characteristics are defined in Table 2.

TABLE II

MAIN CHARACTERISTICS OF THE OBTAINED SET OF CORRELATION COEFFICIENTS.

Characteristic	**Group with positive values**	**Group with negative values**
average correlation coefficient	0.8172	-0.8093
lower quartile	0.7176	-0.9199
upper quartile	0.9122	-0.7161
minimum value	0.5578	-0.9917

Characteristic	**Group with positive values**	**Group with negative values**
maximum value	1.0	-0.577

The distribution diagram of correlation coefficients with positive values is most fully described by the Gauss law with MO equal to 0.82 and COE equal to 0.05. Fig. 3 shows that the set of positive correlation coefficients belongs to the normal distribution.

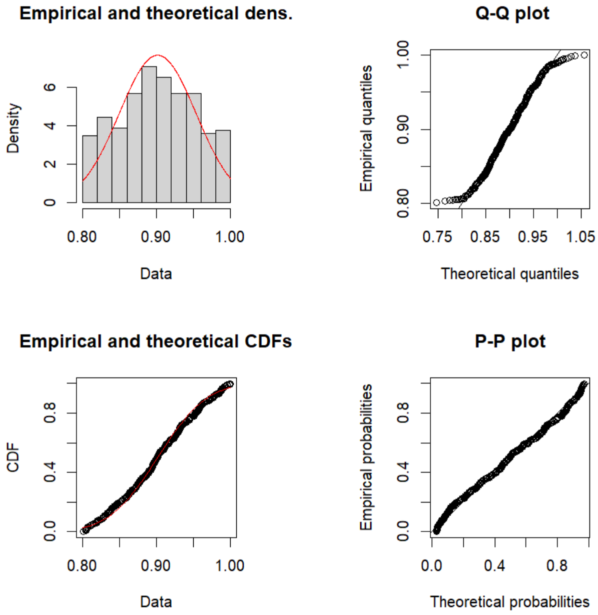

Fig. 3 -Comparative characteristics of the empirical and theoretical distributions for a set with positive correlation coefficients

For this sample, the Kolmogorov-Smirnov criterion was calculated, the P-significance level of which was 0.36, which allows us not to reject the null hypothesis that the sample distribution belongs to the normal one.

When restoring the regression, the values of negative modulo correlation coefficients were taken. The diagram of the distribution of correlation coefficients with negative values is most fully described by the Weibull distribution with MO equal to 0.81 and COE equal to 0.12.

Fig. 4 shows that the set of positive correlation coefficients belongs to the Weibull distribution.

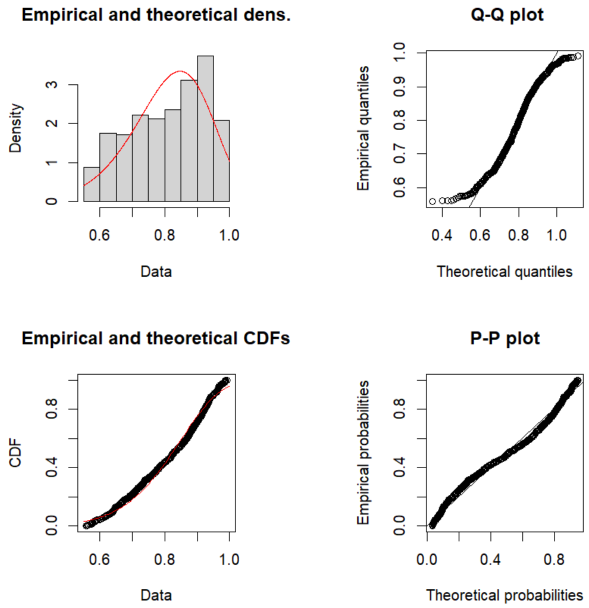

Fig. 4 -Comparative characteristics of empirical and theoretical distribution for a set with negative correlation coefficients

For this sample, the Kolmogorov-Smirnov criterion was calculated with a P-significance level of 0.06, which allows us not to reject the null hypothesis that the sample distribution belongs to the Weibull distribution.

III. DESIGNING THEARCHITECTURE OF A SYSTEM FOR VISUALIZING SUSTAINABLE DEVELOPMENT DATA

The system for visualizing data on sustainable development is a centralized information system designed for conducting sustainability audits of countries and regions on a periodic basis in automatic mode [11].

The system for visualizing data on sustainable development allows you to get an objective assessment of the state of the entire country, as well as individual regions and subjects of the Russian Federation.

The system has three levels of hierarchy:

- federal;
- regional;
- local.

The architecture of the system for visualizing data on sustainable development consists of the following levels [12]:

- collection and primary data processing;
- extracting, converting, and loading data.
- accumulation of data.
- displaying data in data marts.

Below is a more detailed description of each of them.

1) Datacollectionandprimaryprocessing

The first level of the architecture includes data sources called transactional or operational data sources (databases) that are part of the stability analysis system. Transactional databases include data sources that are focused on capturing the results of the system's activities. The requirements for transactional databases have determined their following distinctive features: ability to process data quickly and maintain a high frequency of changes. Transactional databases do an excellent job of handling the barrage of everyday information that must be routinely processed every day.

A set of transactional data sources forms the lower level of the system architecture for visualizing sustainability data.

2) Extracting, converting, andloadingdata

The process of data extraction, transformation, and loading is supported by so-called ETL tools (extraction, transformation, and loading) designed to extract data from various lower — level transactional sources, transform and consolidate them, and upload them to target analytical databases-data warehouses and data marts. At the conversion stage, data redundancy is eliminated, and the necessary calculations and aggregations are performed.

3) Dataaccumulation

The third level of the system architecture for visualizing data on sustainable development includes data sources, which are called data warehouses (from the English Data Warehouse). Data warehouses include data sources focused on storing and analyzing information. Such sources can combine information from several transactional systems and allow you to analyze it in a complex with the use of modern business intelligence software tools.

According to the definition of the founder of the idea of data accumulation, B. Inmon, a data warehouse is a domain-oriented, integrated, uncorrectable, time-dependent collection of data designed to support management decision-making.

Typical features of data warehouses are: relatively rare correct ability of most data, data updatability on a periodic basis, a unified approach to naming and storing data regardless of their organization in the original sources.

The data warehouse, being one of the main parts of the system architecture for visualizing data on sustainable development, acts as the main data source for a comprehensive analysis of all available information [13].

4) Displaying data in data marts

The fourth level of the system architecture for visualizing data on sustainable development includes data sources called data marts, designed for conducting targeted analysis [14]. Data marts are usually built on the basis of information from the data warehouse, but they can also be formed from data taken directly from transactional systems when the data warehouse is not implemented in the organization for some reason.

A. Types of information storage in data marts

According to the type of information storage, storefronts are divided into relational and multidimensional storefronts. Storefronts of the first type are organized in the form of a relational database, where the central table, a fact table intended mainly for storing quantitative information, is linked to reference tables.

Multidimensional storefronts are organized in the form of multidimensional OLAP (Online Analytical Processing)databases, where reference information is presented in the form of dimensions, and quantitative information is presented in the form of indicators. Information in a multidimensional data mart is presented in the most accessible form to end users, which significantly reduces the time required to obtain information for decision -making [15].

From the user's point of view, the difference between data marts and data warehouses is that the data warehouse corresponds to the level of the entire organization, and each showcase usually serves a level no higher than a separate department or process, and sometimes can be created for individual use, differing in a rather narrow target specialization.

The difference between data marts and transactional databases is that the former serve to meet the needs of end users who are not professional programmers: analysts, managers of different levels who solve various tasks. Transactional databases are used for entering and processing primary information, rather than for analyzing it to support decision-making.

The use of data marts, multidimensional and relational, combined with modern business intelligence tools allows you to turn just data into useful information, on the basis of which you can make effective decisions.

The next level of system architecture for visualizing data on the sustainable development of an organization includes modern software tools called Business Intelligence Tools or BI-tools.

BI tools allow you to conduct a comprehensive analysis of information, help you successfully navigate large amounts of data, analyze information, draw objective conclusions based on the analysis and make informed decisions, make forecasts, reducing the risks of making wrong decisions to an acceptable minimum [16, 17].

Data mining tools are used by end users to access information, visualize it, perform multidimensional analysis, and generate both predefined and custom reports [18]. The input information for business analysis is not so much "raw" data from transactional systems, but rather pre-processed data from storage or presented in data marts.

A web application is a traditional type of system architecture for visualizing data on sustainable development. The ability to access information through a familiar Web browser allows you to save on the costs associated with purchasing and maintaining desktop analytical applications for a large number of client locations. The implementation of the Web application allows you to provide analytical information to both users and analysts anywhere in the world connected to the portal via the Internet.

The general architecture of building a system for visualizing data on sustainable development is shown in Fig. 5.

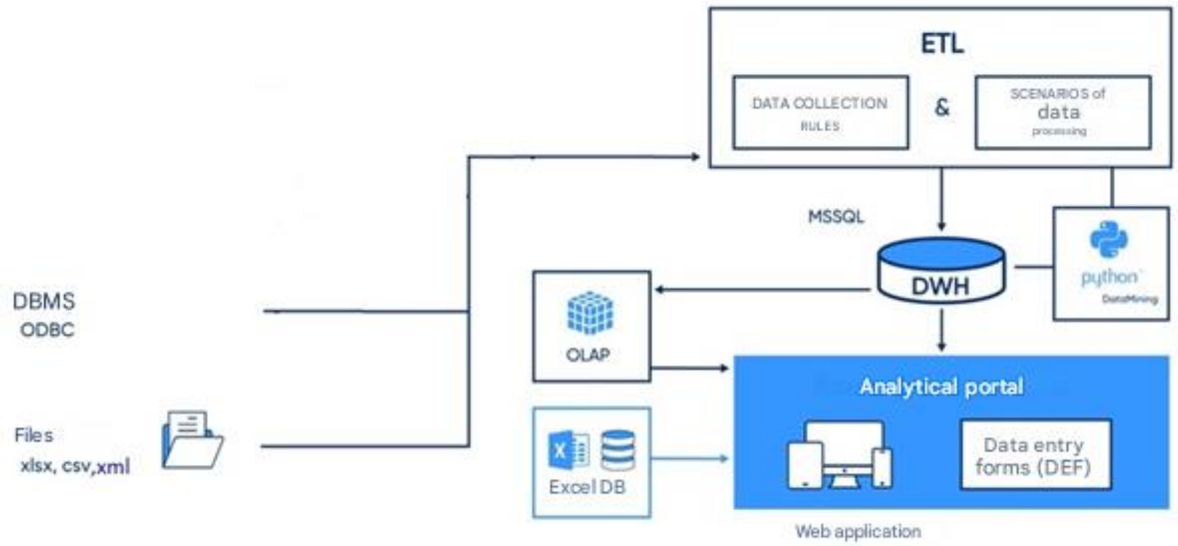

Fig. 5 - Overall system architecture for visualizing sustainability data

B. Designing a database system for visualizing sustainable development

A database is an essential component of any information system, especially in the context of monitoring systems where large amounts of real-time data need to be processed, stored, and accessed. The correct database design determines the speed of query processing, accuracy of analytical reports, and reliability.

For a sustainability data visualization system, a relational database is the most preferred one, поas it provides a strict structure and data integrity.

Relational databases work on the basis of a tabular representation of data, where each table consists of rows and columns. Data in tables can be linked via keys, which allows you to establish relationships between different entities. Queries to these tables are performed using языка SQL (Structured the Structured Query Language (SQL Language), which allows not only retrieving, but also updating, deleting, and inserting data with a high degree of flexibility. In addition, broad support and compatibility with various platforms and tools makes relational databases an ideal choice for system integration and scaling.

To represent the database structure in the system for visualizing data on sustainable development, we will design an Entity-Relationship Diagram. An ER diagram is a graphical representation of entities, their attributes, and the relationships between them. It helps you visualize the database structure, making it clear and intuitive. Fig. 6 shows the database structure.

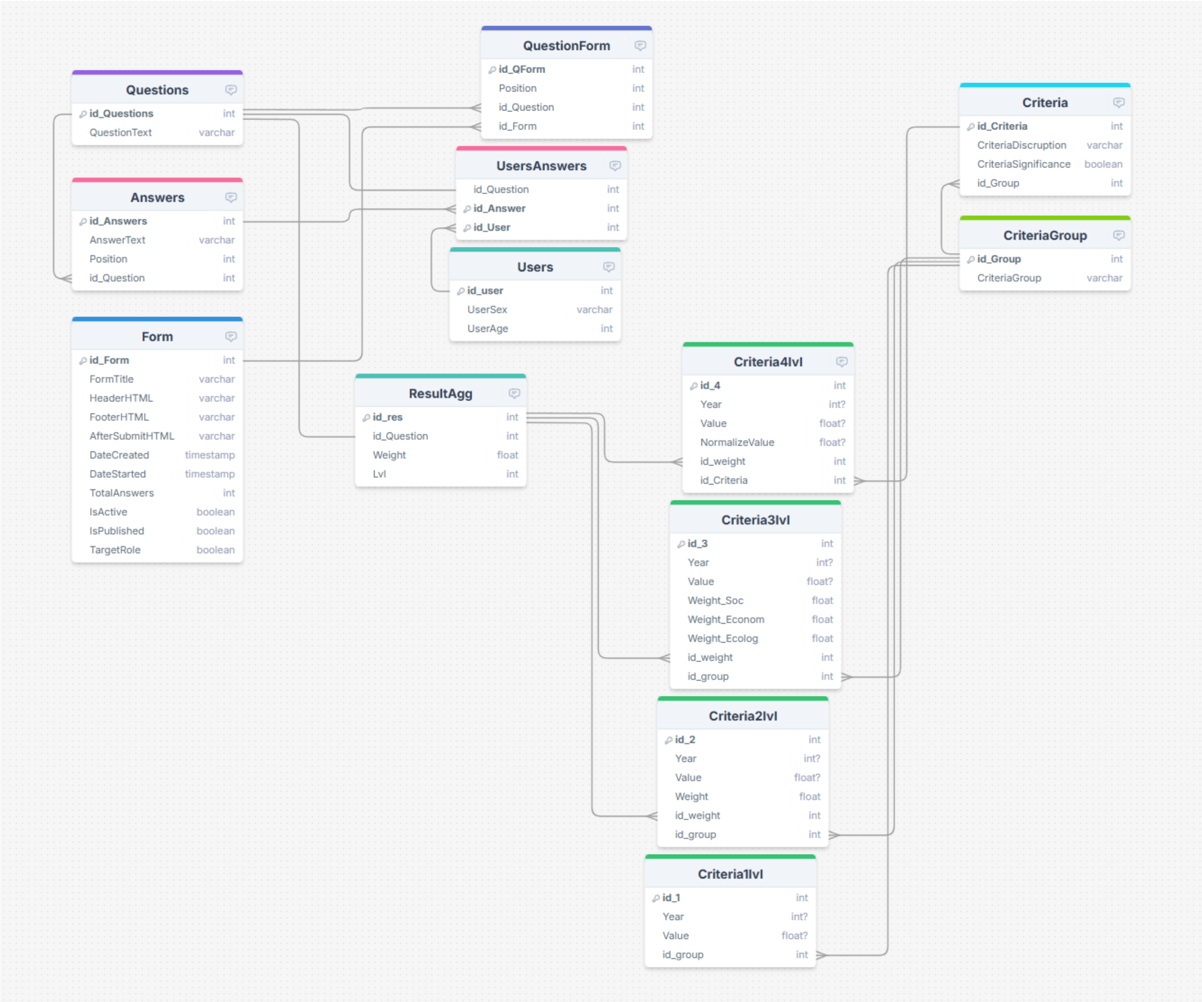

Fig. 6. Database architecture

The diagram is a Boyce-Codd Normal form with one – to-many relationship types. The diagram shows 13 entities: "Questions", "Answers", "Form", "QuestionForm", "UsersAnswers", "Users", "ResultAgg", "Criteria", "CriteriaGroup", "Criteria4lvl", "Criteria3lvl", "Criteria2lvl", "Criteria1lvl". The entities are described below.

The "Questions" entity is responsible for the complete list of questions (Fig. 7). The "id_Questions " attribute is the question ID, and "QuestionText" is the question text.

Questions	
id_Questions	int
QuestionText	varchar

Fig. 7- The "Questions" Entity

The "Answers" entity is responsible for the list of possible answers to all questions from the general list of the sustainability survey (Fig. 8). This entity has information about the identifiers of the question and answer, the text of the answer, and the location of this answer option in the list of answers for a specific question.

Answers	
id_Answers	int
AnswerText	varchar
Position	int
id_Question	int

Fig. 8 - The "Answers" entity

The "Form" entity is responsible for publishing the survey form (Fig. 9). This entity has information about the name of the form, its creation date, publication date and completion date, total number of responses, and status.

Form	
id_Form	int
FormTitle	varchar
HeaderHTML	varchar
FooterHTML	varchar
AfterSubmitHTML	varchar
DateCreated	timestamp
DateStarted	timestamp
TotalAnswers	int
IsActive	boolean
IsPublished	boolean
TargetRole	boolean

Fig. 9 - The "Form" entity

The "QuestionForm" entity is responsible for the list of questions specified in the survey form (Fig. 10). This entity has information about the form and question IDs, and the location of the survey.

QuestionForm	
id_QForm	int
Position	int
id_Question	int
id_Form	int

Fig. 10 - The "QuestionForm" entity

The "UsersAnswers" entity stores each user's answers to all questions (Fig. 11). It consists of a question number, an answer number, and the ID of each user.

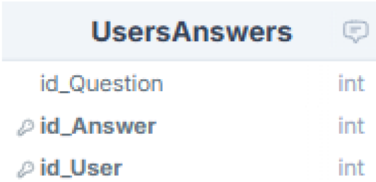

Fig. 11 - The "UsersAnswers" entity

The "Users" entity is responsible for collecting data about the user (Fig. 12). This information may be necessary when analyzing the responses of a particular group of users, broken down by certain characteristics, such as gender and age.

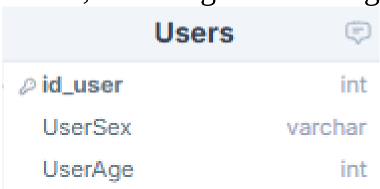

Fig. 12 - The "Users" entity

The "ResultAgg" entity is responsible for the survey results (Fig. 13). It consists of attributes that store information about the weight of each question and the level at which this weight will participate in the calculation.

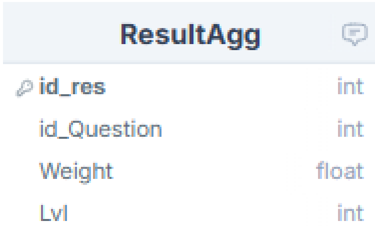

Fig. 13 - Entity "ResultAgg"

The "Criteria" entity is responsible for the complete list of SDG indicators (Fig. 14). The entity contains information about the SDG indicator number, its description, the group to which it belongs, and information about its statistical significance.

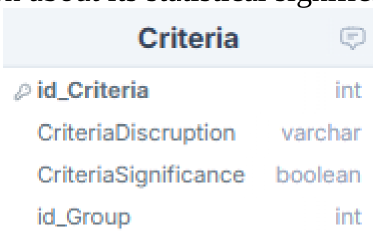

Fig. 14 - The "Criteria" entity

The CriteriaGroup entity is responsible for the list of SDG indicator groups (Fig. 15). Contains the group number and description.

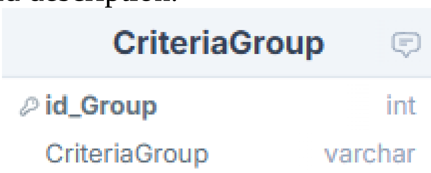

Fig. 15 - The "CriteriaGroup" entity

The "Criteria4lvl" entity is responsible for storing statistically significant SDG indicators for aggregation to the next level (Fig. 16). The entity contains information about the indicator ID, its value, measurement year, normalized value, and weighting factor.

Criteria4lvl	
id_4	int
Year	int?
Value	float?
NormalizeValue	float?
id_weight	int
id_Criteria	int

Fig. 16 - The "Criteria4lvl" entity

The "Criteria3lvl" entity is responsible for storing the values of groups of SDG indicators for aggregation to the next level (Fig. 17). The entity contains information about the group identifier, the aggregated value of the indicators included in it, the year of measurement, as well as the weight coefficients required for calculating the social sustainability indicator, the economic sustainability indicator, and the environmental sustainability indicator.

Criteria3lvl	
id_3	int
Year	int?
Value	float?
Weight_Soc	float
Weight_Econom	float
Weight_Ecolog	float
id_weight	int
id_group	int

Fig. 17 - The "Criteria3lvl" entity

The "Criteria2lvl" entity is responsible for storing the values of the social sustainability indicator, the economic sustainability indicator, and the environmental sustainability indicator (Fig. 18). It contains the indicator ID, year of measurement, value, and weighting factor that are required for aggregation to the next level.

Criteria2lvl	
id_2	int
Year	int?
Value	float?
Weight	float
id_weight	int
id_group	int

Fig. 18 - The "Criteria2lvl" entity

The "Criteria1lvl" entity is responsible for storing a single super criterion (Fig. 19). Contains information about its ID, measurement year, and value.

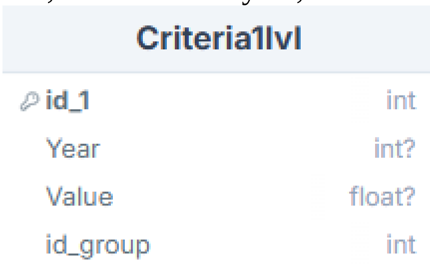

Fig. 19 - The "Criteria1lvl" entity

The presented architecture of the system for visualizing sustainable development data, which will allow for an assessment of sustainability on a federal, regional and local scale, consists of four levels: data collection and primary processing; data extraction, transformation and loading; data accumulation; information presentation in data marts.

The system under development for visualizing sustainable development data will be presented as a Web application that will allow users to analyze it anywhere in the world via the Internet.The presented ER database diagram for building a system for visualizing sustainable development data consists of 9 entities. It will allow you to collect user responses to determine the weighting coefficients for the convolution of sustainability criteria at each level of the graph.

II. CONCLUSIONS

It is proposed to measure sustainability by a single super criterion, the value of which is influenced by sustainability criteria in the social, economic and environmental spheres. The third level includes criteria that determine the solution of 17 tasks to achieve sustainable development goals, which in turn, at the fourth level are 176 criteria. Of these, a group of statistically significant indicators of sustainable development goals was identified. The indicators of this group were normalized and then checked for normality. A correlation matrix of the indicators of sustainable development goals was compiled, which were divided into two groups depending on the nature of the relationship between the indicators - positive or negative. For these groups, the main characteristics were calculated - the average value of the correlation coefficient, the lower quartile, the upper quartile, the minimum and maximum values, and distribution density graphs were constructed. Correlation coefficients between 176 indicators of sustainable development goals were calculated. The main task in this case is to identify statistically significant indicators, normalize them, build a correlation matrix and determine the distribution functions of the values of the correlation coefficients. The following methods were used to solve the tasks: calculation of statistical power under various alternative hypotheses, use of various data normalization methods, computational methods for restoring regression. The main results are: 38% of the indicators from the entire data set turned out to be statistically significant, which were reduced to a single norm by the robust sigmoid function, the correlation coefficients were calculated. In this case, the coefficients with positive values are most fully described by the Gauss law, and the distribution of coefficients with negative values is most fully described by the Weibull distribution, which allows us to determine the mutual influence of SDG indicators for the subsequent construction of a data mart.

A review of existing portal solutions was also conducted, which made it possible to identify a promising direction in the development of portal technologies. The data obtained using the information portal can be applied within the framework of the problem being solved and integrated into the management decision support system.

A generalized architecture of an information system for visualizing data on sustainable development at various scales (federal, regional and local) is presented, which will be presented as a Web application.

A database has been developed, which is part of the system architecture. Its entities allow collecting user responses for expert assessments when determining weighting factors when collapsing sustainable development criteria at all levels of the graph. Also, information on the initial values of indicators, normalized data, aggregated data will be collected and processed in the corresponding entities, which will allow multidimensional analysis in the data mart on a macro and micro scale using various visualization methods.

REFERENCES

[1] Nations, U. (2015). Transforming our world: The 2030 agenda for sustainable development. New York: United Nations, Department of Economic and Social Affairs, 1, 41.J. U. Duncombe, "Infrared navigation—Part I: An assessment of feasibility," IEEE Trans. Electron Devices, vol. ED-11, P. 34-39, Jan. 1959.
[2] United Nations Statistical Commission et al. Annex: global indicator framework for the Sustainable Development Goals and targets of the 2030 Agenda for Sustainable Development. New York: UN Statistical Commission; 2017. – 2018.
[3] Afonichkina E. A. Consistency in achieving sustainable development goals //System Analysis in Design and Management, 2021, vol. 25, no. 3, P. 483-487.
[4] Shevchuk D. S., Il'in I. V. [Digital platform for achieving sustainable development goals]. – 2021. - P. 321-328.
[5] GOST 20886-85. Data organization in data processing systems. Terms and definitions [Text]. - Instead of GOST 20886-75. - Moscow: Standartinform, 2005.
[6] Grzhibovsky AM Correlation analysis // Human Ecology. - 2008. - No. 9. - P. 50-60.
[7] Grzhibovsky AM et al. Sample size for correlation analysis // Marine Medicine. - 2020. - V. 6. - No. 1. - P. 101-106.
[8] Mikhailov V. G., Mezhennaya N. M. About the rate of normal approximation for the distribution of the number of repetitions in a stationary discrete random sequence //PrikladnayaDiskretnayaMatematika. – 2022. – №. 4. – P. 15-21.
[9] Fulcher B. D., Little M. A., Jones N. S. Highly comparative time-series analysis: the empirical structure of time series and their methods // Journal of the Royal Society Interface. – 2013. – T. 10. – No. 83
[10] Royston P. Remark AS R94: A remark on algorithm AS 181: The W-test for normality // Journal of the Royal Statistical Society. Series C (Applied Statistics). – 1995. – T. 44. – No. 4. – P. 547-551.
[11] Information security of open systems: Textbook for universities. In 2 volumes. Volume 2-Security tools in networks / S. V. Zapechnikov, N. G. Miloslavskaya, A. I. Tolstoy, D. V. Ushakov. Moscow: Goryachayaliniya-Telekom, 2008. - 558 p.
[12] Morozova O. A. Integratsiyakorporativnykhinformatsionnykhsistem: uchebnoeposobie [Integration of corporate information systems: a textbook]. Moscow: FinansovyiUniversitet, 2014, 140 p.
[13] W.H. Inmon. Building the Data Warehouse, Fourth Edition // Wiley Publishing, Inc., Indianapolis, Indiana ISBN-13: 978-0-7645-9944-6 - pp.139-157

[14] Shabanova N. Yu.,Efremova O. A., Sharapov A. R. Trends in the development of BIsolutions. // All-Russian scientific and Technical Conference "Problems of Science". Conference materials. Part. 2. Technical Sciences / D. I. Mendeleev Russian State Technical University. - 2019 - P.214-218

[15] Codd, E. F. (1993). Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate. http://www. arborsoft. com/papers/coddTOC. html.

[16] Korotkaya N. V. Modern BI platforms: opportunities and tools //Control systems, information technologies, and mathematical modeling. - 2022. - P. 179-184.

[17] Pak V. V. BI-tools as a base for building an effective internal reporting system / / Accounting, analysis and audit: history, modernity and development prospects. 2019, P. 239-242.

[18] Azeroual O., Theel H. The effects of using business intelligence systems on an excellence management and decision-making process by start-up companies: A case study //arXiv preprint arXiv:1901.10555. – 2019.