DOI 10.31891/2307-5732-2024-333-2-45
УДК 004.623:504.064.36

SLIPCHENKO VOLODYMYR
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
https://orcid.org/0000-0002-3405-0781
e-mail: ddpolytechnic2016@gmail.com
POLIAHUSHKO LIUBOV
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
https://orcid.org/0000-0003-3287-8523
e-mail: liubovpoliagushko@gmail.com
KRUSH OLHA
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
https://orcid.org/0000-0001-8792-2515
e-mail: segroupccte@gmail.com
RUDYK VOLODYMYR
National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute»
https://orcid.org/0009-0004-4774-6579
e-mail: rudykviv@gmail.com

SPECIFICS OF DATABASE STRUCTURE FOR STORING DIFFERENT TYPES OF INFORMATION IN INTEGRATED MONITORING SYSTEMS

Modern integrated monitoring systems of territories and objects, such as countries, regions, cities, districts, enterprises, or even natural resources, deal with a vast amount of diverse information. Indicators characterizing the state of these objects may differ fundamentally and have distinct components, specifics of data processing, and storage. This diversity poses a challenge for standardizing data under a single template, as it requires accommodating the peculiarities of each indicator. The creation of a unified database structure for such systems is a complex and extremely important aspect that can significantly enhance monitoring efficiency and productivity.

This paper discusses the peculiarities of developing a database structure for complex monitoring systems of ecological, economic, energy, and social parameters of territories, where comprehensive analysis requires consideration of data from different fields, emphasizing the importance of creating a flexible database structure that can be easily modified.

A vertical data storage principle is proposed for preserving heterogeneous information, which eliminates the problem of information redundancy and facilitates the development of a universal database structure. The proposed structure remains unchanged when adding new parameters, allowing various experts to work with the system and analyze the environmental state using different indicators, without requiring database restructuring or changes to the software modules utilizing it.

A database structure for storing calculations and formulas employing the vertical data storage principle is developed and presented, aiming to increase data processing speed and reduce time spent on programming new mathematical tools and analysis.

The described database structure is successfully utilized in a comprehensive eco-energetic-economic monitoring system.

Keywords: database design, database structure, integrated monitoring, environment, public health.

СЛІПЧЕНКО ВОЛОДИМИР, ПОЛЯГУШКО ЛЮБОВ, КРУШ ОЛЬГА, РУДИК ВОЛОДИМИР
Національний технічний університет України «Київський політехнічний інститут ім. Ігоря Сікорського»

ОСОБЛИВОСТІ СТРУКТУРИ БАЗИ ДАНИХ ДЛЯ ЗБЕРЕЖЕННЯ РІЗНОТИПНОЇ ІНФОРМАЦІЇ В СИСТЕМАХ КОМПЛЕКСНОГО МОНІТОРИНГУ

У цій роботі розглянуто особливості розробки структури бази даних для систем комплексного моніторингу екологічних, економічних, енергетичних та соціальних параметрів території, де в комплексному аналізі необхідно враховувати дані з різних галузей та є важливим створення гнучкої структури бази даних, що легко можна модифікувати. Запропоновано для збереження різнотипної інформації застосовувати вертикальний принцип розташування даних, що дозволив прибрати проблему надлишковості інформації та розробити універсальну структуру бази даних. Запропонована структура не змінюється при доданні нових параметрів, тому різні експерти можуть працювати з системою та аналізувати стан навколишнього середовища за різними показниками, а зміна набору параметрів не вимагає реструктуризації бази даних та зміни програмних модулів, що її використовує. Розроблено та представлено структуру бази даних для збереження розрахунків та формул, що застосовує вертикальний принцип розташування даних, це дозволяє збільшити швидкість обробки даних та зменшити затрати часу на програмування нового математичного апарату та аналізу. Описана структура бази даних успішно використовується в системі комплексного еко-енерго-економічного моніторингу.

Ключові слова: структура бази даних, комплексний моніторинг, довкілля, здоров'я населення.

Formulation of the problem

One of the important challenges of today is ensuring sustainable development of the country, high quality of life, and population health. Therefore, specialists from Igor Sikorsky Kyiv Polytechnic Institute have developed a Complex Eco-Energy-Economic Monitoring (CEEEM) System [1] for effective decision-making. The CEEEM system allows monitoring of regions, districts, and cities by experts of various profiles, namely: ecologists, energy experts, economists, medical professionals, lawyers, and analysts. Such a set of experts covers all important aspects of the ecological, economic, and social development of the population [2]. Experts can utilize diverse information to conduct object monitoring and provide recommendations for improvement, including information from other spheres.

For example, to calculate damages from excessive emissions, economists use information on emissions of pollutants into the atmosphere, discharges into water bodies, and waste disposal, while for assessing health risks to the population, medical professionals use data on the state of drinking water and soil, presence of harmful substances in the atmosphere, and so on.

Therefore, comprehensive monitoring requires investigation of objects based on various parameters, resulting in database structures (DBs) in such systems being quite cumbersome, hence the issue of optimizing the DB structure of such systems is relevant.

Analysis of the latest research

The structure of a database can significantly impact the range of tasks an automated system is capable of addressing. Typically, automated systems have a narrow focus, and thus the database is designed to accurately describe the subject area. For instance, in [3], a database for modeling river flow is described; in [4], the HydroServer platform is presented, containing sets of spatial-temporal hydrological data; [5] details a database for solid waste management; [6] outlines the structure of an open database concerning global coal and metal mining; [7] describes a database set for working on open rock masses, including databases for air quality, waste management, map operations, and land use; [8] illustrates the database structure for assessing the impact of chemical substances on humans and the environment; [9] discusses the database structure linking pollution levels to children's respiratory illnesses; and [10] presents a database containing information on diseases, climate, and ecological remote sensing data for monitoring animal and human health. All of these are optimized for specific tasks and subject areas. In cases where information from various fields is required, a set of databases is created, each responsible for one area, which are then combined through queries.

However, for comprehensive territory analysis, i.e., one that encompasses all monitoring subsystems [11], task-specific optimization would lead to a large number of specialized tables, and processing them would result in breaking down the complex system into subsystems.

The purpose of this work is to investigate different data layout principles and develop a unified database structure that allows storing heterogeneous specialized information without the need to restructure the database in the CEEEM system.

The data layout principle for preserving different types of information

The proper structure of a database enables the elimination of data redundancy. To achieve this goal, data is partitioned into numerous thematic tables, ensuring each fact is referenced only once.

According to the concept of software flexibility, the database should be designed to accommodate information from various experts without requiring restructuring when expanding the set of formulas or calculation parameters. Realizing such a concept necessitates the provision of a universal structure for primary tables.

Typically, horizontal data layout is employed, where tables consist of a set of fields corresponding to all possible characteristics of real-world objects (the term «horizontal data layout» denotes that all attribute values for a single object are contained within one tuple, which represents one horizontal row in the table). For example, to store data regarding the macro-component composition of groundwater, a table may include approximately the following fields: object code, observation point number, odor, color, carbonates, hydrocarbonates, chlorides, sulfates, nitrites, nitrates, hardness, magnesium, calcium, iron, ammonium, sodium, polyphosphates, fluorides [12]. This list of characteristics may vary depending on the need for a more or less detailed examination of water composition.

Having a separate field for each characteristic facilitates data entry and processing. However, if the set of characteristics needs to be expanded, such a data model requires a change in the table structure. Conversely, a reduction in the parameter set leads to empty cells, resulting in suboptimal memory utilization. It is evident that such an approach is ideal for a narrow domain and processing homogeneous information but is not suitable for comprehensive multifaceted analysis.

If the database is intended to store different types of information, it is necessary to apply the vertical data layout principle (the term «vertical data layout» means that all indicator values for one object are contained within one column of the table, i.e., vertically). The principle of constructing such a structure can be described by the following formula:

$$T_{object} + T_{object\ indicators} + T_{indicators} = T_{full\ description}, \quad (1)$$

where T_{object} – a table for storing general information about the object, regardless of its type;

$T_{object\ indicators}$ – a table containing a list of values only for those indicators that are important for a specific object;

$T_{indicators}$ – a table containing a list of ALL possible indicators, essentially serving as a reference for indicators;

$T_{full\ description}$ – a virtual table that does not physically exist, representing the complete description of the object without redundant data, which can be obtained by selecting data from the three tables.

In the CEEEMS system, this principle is implemented in the part of the database that stores measurement results, calculation results, and formulas.

The database structure for storing measurement results

For the given data in the database, tables have been created: «POI Objects» (containing general information about monitoring objects on the map), «Elements» (listing pollutants and parameters of the investigated environment), «Emissions on the map» (containing information about measurement results at specific points on the map), and «Environment» (listing components of the surrounding environment). The picture (Fig. 1) presents the corresponding fragment of the database.

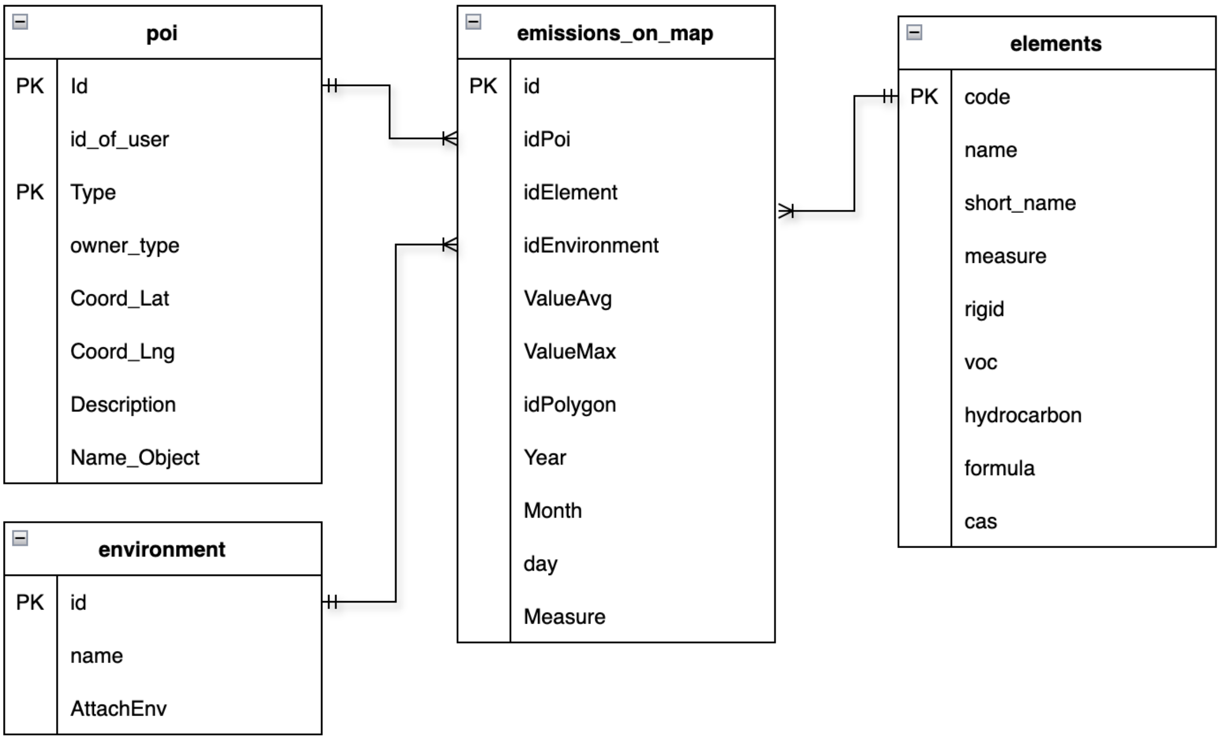

Fig. 1. Scheme of a part of the database for saving measurements

According to the approach described above: T_{object} = «POI Objects», $T_{indicators}$ = «Elements», a $T_{object\ indicators}$ = «Emissions on the map».

Table 1 presents a fragment of the «POI Objects» table. The fragment contains information about two objects of different types, belonging to different monitoring subsystems. The Pump belongs to the water environment monitoring subsystem, while «SAVEDNIPRO_1376» represents an air quality monitoring station, specifically the SaveEcoSensor 3.0, and belongs to the atmospheric air monitoring subsystem.

Table 1

A fragment of the table «POI Objects»

ID	Object type	Latitude	Longitude	Address	Name
1417	Pump	50.3659652205212	30.4567515850067	Akademika Hlushkova Ave., 39	Pump on Akademika Hlushkova Ave., 39
1147	Place of information collection	50.421359	30.514859	Vododymyro-Lybidska Str., 25	SAVEDNIPRO_1376

The CEEEMS receives data on atmospheric air pollution from the SaveEcobot web system according to two indicators: fine dust fraction 2.5 (PM2.5) and fine dust fraction 10 (PM10).

Table 2 presents a fragment of the «Emissions on the map» table, which contains air quality indicators at point SAVEDNIPRO_1376.

Table 2

A fragment of the table «Emissions on the map» with air quality indicators

ID	POI Object	Indicator	Environment	Value	Year	Month	Day	Unit
1	SAVEDNIPRO_1376	PM2.5	Air	18,6	2021	11	5	$\mu g / m^3$
2	SAVEDNIPRO_1376	PM10	Air	32,5	2021	11	5	$\mu g / m^3$

The structure of the table is also suitable for storing information on the contamination of drinking water from pumps according to the following indicators: turbidity, color, total hardness, manganese, total iron, ammonium, smell at 20°, taste and aftertaste.

Table 3 presents a fragment of the table «Emissions on the map» with data on the quality of water in the pumping station on Akademika Hlushkova Ave., 39. Table 4 presents a fragment of the table «Emissions on the map» with data on soil indicators.

Table 3

A fragment of the table «Emissions on the map» with data on water quality

ID	POI Object	Indicator	Environment	Value	Year	Month	Day	Unit
1	Pump on Akademika Hlushkova Ave., 39	Turbidity	Drinking water	0	2021	06	25	mg / dm^3
2	Pump on Akademika Hlushkova Ave., 39	Coloration	Drinking water	17.8	2021	06	25	Degree
3	Pump on Akademika Hlushkova Ave., 39	Total hardness	Drinking water	1.5	2021	06	25	mg / dm^3
4	Pump on Akademika Hlushkova Ave., 39	Manganese	Drinking water	0.011	2021	06	25	mg / dm^3
5	Pump on Akademika Hlushkova Ave., 39	Total iron	Drinking water	0.15	2021	06	25	mg / dm^3
6	Pump on Akademika Hlushkova Ave., 39	Ammonium	Drinking water	0.07	2021	06	25	mg / dm^3
7	Pump on Akademika Hlushkova Ave., 39	Smell at 20°	Drinking water	0	2021	06	25	Point
8	Pump on Akademika Hlushkova Ave., 39	Taste and aftertaste	Drinking water	0	2021	06	25	

Table 4

A fragment of the table «Emissions on the map» with data on soil indicators

ID	POI Object	Indicator	Environment	Value	Year	Month	Day	Unit
1	Kyiv region	Reaction of the soil solution	Soil	5.6	2021	06	17	pH
2	Kyiv region	Humus	Soil	2.1	2021	06	17	%
3	Kyiv region	Moving phosphorus compounds	Soil	101	2021	06	17	mg/kg
4	Kyiv region	Moving potassium compounds	Soil	81	2021	06	17	mg/kg

In the «Emissions on map» table, the «Indicator» column is filled from the «Elements» table, a fragment of which is presented in Table 5.

Table 5

A fragment of the table «Elements»

ID	Name	Unit	Solubility	Formula
1	fine dust PM2.5	$\mu g/m^3$	-	
2	fine dust PM10	$\mu g/m^3$	-	
3	Turbidity	mg/dm^3	-	
4	Coloration	Degree	-	
5	Total hardness	mg/dm^3	-	
6	Manganese	mg/dm^3	+	MnO
7	Total iron	mg/dm^3	-	
8	Ammonium	mg/dm^3	+	NH_4
9	Smell at 20°	Point		
10	Taste and aftertaste			

Medical statistics and energy indicators are stored in the same format. As seen, the number of indicators characterizing the monitoring object does not affect the database structure. Therefore, various experts can work with the system and analyze the state of the environment using different indicators, and changing the set of parameters does not require restructuring the database or changing the software modules that use it.

The database structure for storing calculation results and formulas.

Following a similar principle, a part of the database is organized to store information about calculation results and the formulas themselves. It includes the following tables: «Formulas», «Formula compound», «Calculation results» and «Parameter's value». The picture (Fig. 2) shows a fragment of the database intended for storing formulas and calculation results.

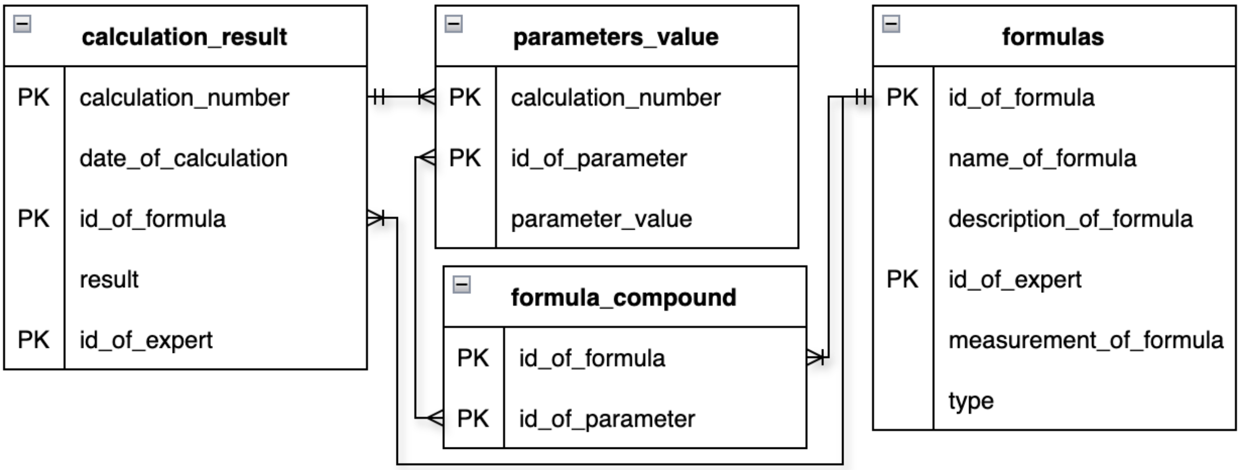

Fig. 2. Scheme of a part of the database for storing calculation results and formulas

According to the approach described above: for the part of the database intended for storing formulas T_{object} = «Formulas», $T_{indicators}$ = «Formula parameters», a $T_{object\ indicators}$ = «Formula compound»; for the part of the database intended for saving calculation results: T_{object} = «Formulas», $T_{indicators}$ = «Formula parameters», a $T_{object\ indicators}$ = «Formula compound».

The «Formula compound» table (Table 7) is intended for assembling the set of parameters necessary to compute a value according to a formula. The «Formulas» table (Table 6) contains a list of formulas available in the program for conducting calculations. Since a formula can be a parameter itself, the «Formula parameters» table is essentially the same as the «Formulas» table, but the value of the «Type» field determines whether this entry is a regular parameter or a formula itself. For example, the soil pollution coefficient C_s is determined by the formula:

$$C_s = \frac{V_p}{T_{sl} \times A_{pl} \times I_c}, \quad (2)$$

where C_s – soil pollution coefficient,
V_p – volume of pollutant,
T_{sl} – thickness of soil layer,
A_{pl} – the area of polluted land,
I_c – correction index for pollution cleanup costs depending on the depth of substance penetration.

Almost all formula parameters are common parameters that are measured, not calculated, but the volume of the pollutant V_p is itself a formula:

$$V_p = \frac{M_p}{D_p}, \quad (3)$$

where M_p – mass of pollutant,
D_p – relative density of the pollutant.

Taking formulas (2) and (3) into account, the «Formulas» table will look like Table 6, and the «Formula compound» table will look like Table 7.

It can be concluded that the presented database structure allows for storing information about formulas related to any environmental monitoring subsystem. This means that without changing the structure, a system using such a database can provide mathematical tools for any expert working in the monitoring field.

Table 6

Formulas

ID	Parameter or formula name	Description	Unit	Type
1	Soil pollution coefficient	C_s		formula
2	Volume of pollutant	V_p	m^3	formula
3	Thickness of the soil layer	T_{sl}	m	parameter
4	Area of polluted land	A_{pl}	m^2	parameter
5	Correction index for pollution cleanup costs	I_c		parameter
6	Mass of pollutant	M_p	t	parameter
7	Relative density of the pollutant	D_p	t/m^3	parameter

Table 7

Formula compound

Formula ID	Parameter ID
1	2
1	3
1	4
1	5
2	6
2	7

Table 8

A fragment of the table «Calculation results» with data on energy indicators

ID	POI Object	Formula	Result	Description	Date
1	Preschool educational institution «Berehynia»	ECV_y	48941.04	Electricity consumption volumes	16.03.2021
2	Preschool educational institution «Berehynia»	EOE	48941.04	Expenditure on electricity	16.03.2021
3	Preschool educational institution «Berehynia»	HECV	662.055	Heat energy consumption volumes	16.03.2021
4	Preschool educational institution «Berehynia»	EOHEON	662.055	Expenditure on heat energy for own needs	16.03.2021
5	Preschool educational institution «Berehynia»	WCV_y	874.78	Water consumption volumes	16.03.2021
6	Preschool educational institution «Berehynia»	EOW	874.78	Expenditure on water	16.03.2021
7	Preschool educational institution «Berehynia»	ECV_ay	4078.42	Average yearly electricity consumption volumes	16.03.2021
8	Preschool educational institution «Berehynia»	WCV_ay	72.89834	Average yearly water consumption volumes	16.03.2021
9	Preschool educational institution «Berehynia»	EOE_ay	4078.42	Average yearly expenditure on electricity	16.03.2021
10	Preschool educational institution «Berehynia»	EOW_ay	72.89834	Average yearly expenditure on water	16.03.2021
11	Preschool educational institution «Berehynia»	ECV_am	2358.09	Average monthly electricity consumption volumes	16.03.2021
12	Preschool educational institution «Berehynia»	WCV__am	42.15601	Average monthly water consumption volumes	16.03.2021
13	Preschool educational institution «Berehynia»	EOE_am	2358.09	Average monthly expenditure on electricity	16.03.2021
14	Preschool educational institution «Berehynia»	EOW_am	42.15601	Average monthly expenditure on water	16.03.2021
15	Preschool educational institution «Berehynia»	HECV_ay	55.17125	Average yearly heat energy consumption volumes	16.03.2021
16	Preschool educational institution «Berehynia»	HECV_am	32.0841	Average monthly heat energy consumption volumes	16.03.2021
17	Preschool educational institution «Berehynia»	EOHE_ay	32.0841	Average monthly expenditure on heat energy	16.03.2021
18	Preschool educational institution «Berehynia»	EOHE_am	55.17125	Average yearly expenditure on heat energy	16.03.2021

Similarly, another part of the database, intended for storing calculation results (Table 8), is organized in the same way. These results can subsequently be used for analyzing the state of the territory.

Such organization of formulas allows for changing the formula and parameter set with minimal impact on the program code.

Conclusions

The proposed application of the vertical data arrangement principle, described by formula (1), has addressed the issue of data redundancy and facilitated the development of a universal database structure.

The proposed database structure for storing diverse information remains unchanged when adding new parameters. Therefore, different experts can work with the system and analyze the environmental status using various indicators. Additionally, modifying the parameter set does not necessitate restructuring the database or changing the software modules that utilize it.

The suggested structure of the database for storing calculations and formulas employs the vertical data arrangement principle, enhancing data processing speed and reducing the time required for programming new mathematical apparatus and analysis, in contrast to the traditional approach to database design.

The proposed database structure enables comprehensive analysis of the territory's status without dealing with numerous narrowly focused tables. It is successfully utilized in a complex eco-energy-economic monitoring system.

References

1. Slipchenko V., Poliahushko L., Krush O. The system of complex eco-energy-economic monitoring to optimize management decisions (region, district and city). *Visnik of the Volodymyr Dahl East Ukrainian National University*. 2021. №4(268), P. 13-20. DOI: 10.33216/1998-7927-2021-268-4-13-20

2. Slipchenko V., Poliahushko L., Krush O. Developing an automated system for collecting various information in the system of complex eco-energy-economic monitoring. *Technology Audit and Production Reserves.* 2022. № 3/2 (65), P. 11–18. DOI: 10.15587/2706-5448.2022.259069

3. Erik Bollen, Brianna R. Pagán, Bart Kuijpers, Stijn Van Hoey, Nele Desmet, Rik Hendrix, Jef Dams, Piet Seuntjens; A database system for querying of river networks: facilitating monitoring and prediction applications. *Water Supply* 1 March 2022. 22 (3). P. 2832–2846. DOI : 10.2166/ws.2021.433

4. Horsburgh, Jeffery & Tarboton, David & Schreuders, Kimberly & Maidment, David & Zaslavsky, Ilya & Valentine, David. HydroServer: A Platform for Publishing Space-Time Hydrologic Datasets. *AWRA 2010 Spring Specialty Conference - GIS and Water Resources.* March 2010. P. 1-6.

5. Wang, J. & Li, C. & Shi, Z. & Dong, J. & Feng, T.. Illegal Municipal Solid Waste Monitoring System Based On Android App. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. V-3-2020. P. 781-786. DOI : 10.5194/isprs-annals-V-3-2020-781-2020.

6. An open database on global coal and metal mine production / S. Jasansky et al. *Scientific Data*. 2023. Vol. 10, no. 1. DOI : 10.1038/s41597-023-01965-y

7. Design of a GIS Database for Surface Mining / G. M. Vangu et al. Journal of Applied Engineering *Sciences*. 2023. Vol. 13, no. 2. P. 289–296. DOI : 10.2478/jaes-2023-0037

8. A harmonized chemical monitoring database for support of exposure assessments / K. K. Isaacs et al. *Scientific Data*. 2022. Vol. 9, no. 1. DOI : 10.1038/s41597-022-01365-8

9. Dunea, Daniel & Iordache, Stefania & Oprea, Mihaela & Savu, Tom & Pohoata, Alin & Lungu, Emil. A Relational Database Structure for Linking Air Pollution Levels with Children's Respiratory Illnesses. *Bulletin USAMV series Agriculture.* 2014. 71. P. 205-213. DOI : 10.15835/buasvmcn-agr:.

10. A Web Geographic Information System to share data and explorative analysis tools: The application to West Nile disease in the Mediterranean basin / L. Savini et al. *PLOS ONE*. 2018. Vol. 13, no. 6. P. e0196429. DOI : 10.1371/journal.pone.0196429

11. Ecological monitoring of the environment. Ministry of Ecology and Natural Resources of Ukraine. URL: https://mepr.gov.ua/diyalnist/napryamky/ekologichnyj-monitoryng/ekologichnyj-monitoryng-dovkillya/ (Accessed: 15.01.2024).

12. On the approval of the Procedure for State Water Monitoring. Resolution of the Cabinet of Ministers of Ukraine № 758 of 19.09.2018. URL: https://zakon.rada.gov.ua/laws/show/758-2018-%D0%BF#Text (Accessed: 15.01.2024).