Mechanisms of polymorphic systematization of bioecological data within the BaikalIntelli platform for organizing computational models of population dynamics

V S Kedrin, A V Arguchintsev and I M Dobrinets
Department of Computational Mathematics and Optimization, Irkutsk State University, 1 K. Marks St., 664003 Irkutsk, Russia
E-mail: kedrinvs@isu.ru

Abstract. A conceptual model of the BaikalIntelli platform for universal storage, input and processing of arbitrary spatio-temporal databases of bioecological parameters is proposed. Information technologies that allow implementing the basic functionality of the universal metamechanism of data systematization are considered. The platform mechanism has been tested on the example of entering quantitative data of different types of oligochaetes, one of the mass groups of invertebrates of Lake Baikal. Web interface for the database “Spatio-temporal variability of oligochaete populations in the area of industrial wastewater discharge of the Baikal pulp and paper mill” is developed on the base of JavaScript.

1. Introduction
The project of the BaikalIntelli research platform aims to develop a universal flexible information and analytical tools that can be easily adapted for heterogeneous data of any size, structure and geographical reference (pollution by arbitrary components; hydrological and meteorological parameters; data on the species, quantitative and age composition of biological populations, etc.). The unique format of the platform will allow you to comprehensively focus on the possibility of multiple system post-processing and data analysis for building mathematical models for assessing the anthropogenic impact and environmental situation of the lake in its various sections, predicting situations, including in cases of accidental releases, natural and man-made disasters, etc. It is intended to create accessible web interfaces for users of any level who are not specialists in the field of information technology. The term Intelli is proposed by the authors with due regard to the requirements for modern aggregate business analysis systems (https://en.wikipedia.org/wiki/Business_intelligence). The platform being developed should provide solutions to the following General tasks:

- analytics, including analysis of heterogeneous large-scale data, statistical analysis, etc.;
- collaboration of specialists of different branches;
- knowledge management;
- measurement and reporting, i.e. getting the necessary selected and summary data for the specified criteria.
2. Materials and methods

The goal of the BaikalIntelli platform is to create a polymorphic universal data systematization mechanism that allows storing, entering and processing arbitrary spatio-temporal databases with different table structures and details to implement the task of forming computational algorithms for calculating mathematical models of population dynamics using optimal control methods. Figure 1 shows the conceptual ER model of the database.

![Entity relations model of the BaikalIntelli platform database.](image)

**Figure 1.** Entity relations model of the BaikalIntelli platform database.

The conceptual mechanism of the BaikalIntelli platform implements the following requirements:

- organization of data input in the context of global directions and nested sections;
- tree architecture of sections within the direction by organizing a hierarchical structure of data storage;
- arbitrary table template schemes within the scientific section, described using JSON format;
- entering an arbitrary number of tables within the scientific section using a universal data storage mechanism in JSON format;
• binding to the geographical location within the section and the data collection table;
• the ability to describe the section/table using MarkDown format;
• a system for accessing registered users to data within the section;
• a universal data storage mechanism in JSON format;
• a distributed REST-exchange model between client-side and server-side of the application.

The asynchronous web framework Django 3.0 in Python [1], which uses the Model-View-Controller (MVC) design pattern, was chosen as the basis for the basic functionality of the universal metamechanism for data systematization [2]. The design pattern involves a separate description of the data model using the ORM (Object-Relational Mapping) that associates a physical database with an object-relational mapping, allowing the use of more complex abstractions in the server-side application logic, in particular, worked in the platform BaikalIntelli ORM allows:

• to form a clear description of the database schema presented in terms of Python;
• to automate the creation of SQL queries, which eliminates the need to use the language for describing the database structure (Data Definition Language) and the data Manipulation Language (Data Manipulation Language) when designing the database and changing its schema, respectively;
• to use elements of the application’s object logic, i.e. classes, objects, attributes, and methods, rather than elements of the relational data model;
• to use object logic to create a complex hierarchical data organization with inheritance mapping and data table composition;
• to create new SQL queries when migrating to another database management system, because the low-level ORM driver is responsible for this;
• to optimize testing of a significant amount of program code, which is often monotonous and error-prone;
• to isolate program code from data storage details.

As a physical DBMS that interacts with the server logic of the application based on the ORM software technology, the PostgreSQL 12.2. DBMS is the most efficient and developed freely distributed object-relational database management system (ORDBMS). Currently, PostgreSQL 12.2 is a real alternative to commercial databases in terms of reliability, performance, and extensibility. It is also worth noting the wide scalability of the latest versions of the PostgreSQL DBMS due to mechanisms:

• Manage buffers and cache data to maintain efficient use of allocated memory resources.
• Tablespaces for managing data storage at the level of logical DBMS objects, such as databases, schemas, tables, and indexes.
• Data integrity due to a system of locks at the lower level, which allows you to maintain a high level of reliability.
• Supports search indexes B-tree, hash, R-tree, GiST (generalized search tree), partial indexes, and expressive indexes.

An entity relations model of the BaikalIntelli platform has been developed that provides a server-side logic of user database storage mechanisms, which consists of the following models implemented at the ORM level of the application architecture:

(1) “ScienceTheme” – a data model designed to generate a list and describe global scientific research directions (projects).
(2) “LocalScienceTheme” – a hierarchical data model designed to describe local scientific sections within the project associated with the research site, and allows to organize a logical tree-like structure of user databases.

(3) “SchemaTableSection” – a data model designed for a universal description of the schema for storing table data using JSON format. You can see the actual representation of the stored code in figure 2.

(4) “DataScienceCollection” – a data model designed to describe the tables of user databases of research data within a given table data storage scheme using the “SchemaTableSection” model.

(5) “JSONCollection” – a data model designed for storing individual records of user database tables within a specified table data storage scheme using the JSON format.

Figure 2. Storing the table schema structure in JSON format.

Within the framework of the “LocalScienceTheme” model, a mechanism for creating a tree-like list of records was implemented based on the mptt (modified-preorder-tree-traversal) hierarchical data organization technology at the data model level [3]. This mechanism allows:

- to create a recursive hierarchical list at the parent–child relationship level;
to recursively traverse the constructed data tree;
• to add nodes to the data tree;
• to update nodes in the data tree;
• to provide a cascading delete of the branches of the tree and the dependent data.

Separately, the “description” attributes are allocated for describing stored data, which allow you to describe the stored data using the lightweight markup language Markdown. This format allows you to generate high-quality formatted html page markup in the client web interfaces for displaying user databases in data table description blocks in a simplified format.

At the level of the “SchemaTablesection” and “JSONCollection” models, a universal data storage mechanism is implemented using the JSON format, which allows you to hierarchically describe data in the form of two data structures (see figure 2):

• A collection of key/value pairs that can be interpreted as a structure or dictionary at the level of programming languages.
• An ordered list of values that can be interpreted as an array, list, or sequence at the level of programming languages.

3. Results and discussion
Thus, in the first stage of development of the research platform BaikalIntelli, taking into account the technologies described above, the following steps were implemented:

(1) a web resource (http://194.190.232.169) based on Ubuntu LTS 18.04.4 provided PostgreSQL 12.2;
(2) designed and implemented server-side logic universal metamerism systematization of data in accordance with a conceptual ER-model (figure 1) using the web framework Django 3.0;
(3) the developed algorithm is universal parser data files in Excel format (in the programming language Python);
(4) the User database “Spatio-temporal variability of oligochaete populations in the area of industrial wastewater discharge of the Baikal pulp and paper mill” was designed and filled in with a breakdown of data collection sites (based on the BaikalIntelli platform);
(5) implemented data filtering mechanisms based on ORM technology;
(6) the initial tabular web interface of the client part of the output of the user database “Spatio-temporal variability of oligochaete populations in the area of industrial wastewater discharge of the Baikal pulp and paper mill” was implemented (figure 3) (Bootstrap user interface library, JavaScript programming language).

4. Applications for observational data
The user database “Spatio-temporal variability of oligochaete populations in the area of industrial treated wastewater of the Baikal pulp and paper mill” implemented on the basis of the BaikalIntelli platform [4] contains information about hydrobionts obtained in the hydrobiological monitoring system of lake Baikal. Observations of the state of the biota of Lake Baikal were organized in the 20-30s of the last century by prof. M.M. Kozhov, and subsequently continued and improved by prof. O. M. Kozhova, Director of the research Institute of biology of Irkutsk State University. Long-term studies of the Lake Baikal benthic were carried out in the laboratory of hydrobiology of the research Institute of biology of Irkutsk State University, headed by PhD of Biology L.A. Izhboldina. Sections of the user database include materials that were transferred
and preserved by one of its developers – PhD of Biology L.S. Kravtsova, who took part in the collection and processing of samples from her student years, and then led field research (until 1998) as the head of expeditions. Quantitative sampling was carried out with the participation of divers V.A. Murakhveri, V.I. Maksimov, V.A. Gombreich, B.I. Dmitriev on research vessels of the Institute of Biology: “M.M. Kozhov”, “Naturalist”, “Cosmos”. One set of data provides quantitative indicators of each invertebrate species in samples: number-in specimens, biomass, i.e. air-raw weight – in mg. In some data series, quantitative indicators are indicated, which are given in terms of 1 m² of the bottom of lake Baikal. Quantitative samples were collected on stony ground from a recording frame, with an area of 0.1 m², and from silty and sandy bottom sediments – with a Petersen dredger, with a capture area of 0.025 m². Coefficients of 10 or 40, corresponding to the type of bottom sediments, were used to recalculate the Lake Baikal bottom by 1 m². The section "Oligochaetes" includes the materials of PhD of Biology T.V. Akinshin and I.V. Lezinskaya, which identified the species of small-scale worms, determined their number and biomass. Data on oligochaetes from paper to electronic form were translated by M. A. Voylo, clarified and verified by PhD of Biology L.S. Kravtsova. The data contains information about the abundance and biomass of different species of oligochaetes in the Utulik – Hara-Murin rivers area near the East coast of Lake Baikal in the summer and in the subglacial period in March in different years. The designed mechanisms of BaikalIntelli are used to ensure reliable storage and rapid extraction of data from observations of oligochaete populations using different hierarchical systematization from the point of view of factor characteristics and imposed conditions for selecting information. The database interface is aimed at obtaining data using the global Internet (REST API) in csv format in the framework of solving data analysis problems and providing calculations of computational models in the field of mathematical modeling of population dynamics for the formation of specific criteria for assessing the anthropogenic impact on the environmental situation of the Lake Baikal. Table data schemas describe the following data elements: 1) date; 2) Lake; 3) Area; 4) Section; 5) Station; 6) Site; 7) Characteristic (distance); 8) Depth, m; 9) bottom sediments; 10) Collection tool; 11) View; 12) the Number of...
copies in the sample; 13) Biomass, mg in the sample. Further development of the project offers its scaling and development within:

1) Services for solving urgent problems of environmental monitoring of Lake Baikal.

2) Areas related to the construction of complex mathematical models using methods of computational mathematics and statistics, as well as using modern approaches in the field of machine learning.

3) Animation of dynamic mathematical models of pollution to assess the anthropogenic impact on the lake Baikal.

4) Modernization of the platform in the direction from databases to intellectualized knowledge bases.

5) Highlighting the educational direction of the platform using machine learning approaches.

6) Promotion of environmental knowledge in the framework of a comprehensive program for the conservation of nature and the water ecosystem of the Lake Baikal.

Acknowledgments

The reported study was funded by RFBR and the Government of the Irkutsk Region, project number 20-41-385002.

References

[1] Django web framework [website] www.djangoproject.com
[2] Kedrin V S and Kuzmin O V 2011 Application of software architecture Model-View-Controller for building a system of visual design and simulation of algorithms Nauchno-Tekhnicheskiy Vestnik Sankt-Peterburgskogo Gosudarstvennogo Universiteta Informatsionnykh Tekhnologii, Mekhaniki I Optiki [Scientific and Technical Bulletin of St. Petersburg State University of Information Technologies, Mechanics and Optics] 4 36-42
[3] Van Tulder G 2003 Storing hierarchical data in a database https://www.sitepoint.com/hierarchical-data-database/
[4] Arguchintsev A V, Kedrin V S, Kravtsova L S, Dobrinets I M and Glyzina (Voylo) M A Metatechnology of designing a hierarchical modifiable-intersecting database for remote control of bioecological parameters 2020 Limnology and Freshwater Biology 4 641-2