

Formation of Reports for Model-Driven Data Consolidation System*

Aleksei Korobko^[0000-0002-6227-1362]

Institute of Computational Modeling of the Siberian Branch
of the Russian Academy of Sciences, 50/44 Akademgorodok, Krasnoyarsk, 660036, Russia
agok@icm.krasn.ru

Abstract. A platform for constructing model-driven systems was developed by the staff of ICM SB RAS to support the consolidation, storage and analytical processing of monitoring data for the state of natural and technogenic objects, medical data, scientific research data, etc. The basis for the creation of the platform was an original approach to the construction of model-driven systems, a meta-meta model and a set of algorithms. Algorithms are responsible for navigating between models, creating structures in the database, building a user interface, etc. One of the problems that have not been solved within the framework of the platform for building data consolidation systems is the automation of the creation of analytical reports. The existing solution makes it difficult for users to independently conduct analytical experiments, as it involves contacting the administrator or mastering the skills of working with SQL queries. The task of developing tools for the native formation of analytical queries to data is urgent. The article proposes an algorithm for generating queries and a constructor of research queries that implements this algorithm.

Keywords: Web-system, Ad-hoc Data Consolidation, Model-driven Development, Dynamic User Interface, Metadata, Report.

1 Introduction

The staff of ICM SB RAS have developed many systems focused on data collection. With the help of the system for consolidation of monitoring data of the regional center for monitoring and forecasting emergencies, it is possible to keep under observation the state of natural and technogenic objects, for example, the water level in the rivers and the number of accidents on the roads [1]. A consolidation system has been developed for scientific organizations which allows recording the results of scientific work: publication activity of employees, patent activity, etc. [2, 3]. Much work has been done to facilitate the collection and storage of research results [4]. Systems have been developed for collecting medical questionnaire information, environmental

* Copyright c 2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

monitoring data of the Krasnoyarsk reservoir [5] and data for assessing the ecological state of soils in the Krasnoyarsk Region.

The main feature of the constructed data consolidation systems is the use of the original modification of the model-driven approach in their development (MDD – model-driven development) [6, 7]. MDD uses models as a central component of the software development process. All the models are divided into levels of abstraction, the transition from the models of one level to the models of another level is regulated. The result of the development is the program code of the system. The modification consists in using the tools of the data consolidation system to edit the control model and instantly change the behavior of the system in response to the change in the model [8]. The author's implementation of the model-driven approach allows the dynamic evolution of the created data consolidation systems by the users themselves.

Experience (more than five years) in creating model-driven systems for consolidating and storing data, as well as systematizing the accumulated knowledge, made it possible to develop a universal meta-metamodel of the data collection process. The meta-metamodel describes the main classes of entities and relationships involved in the data collection process. They are used to build control models for specific application systems. Also, a set of algorithms for automatic generation of the applied models and a system interface based on the control model was created and tested. Together with the original approach to the construction of model-driven systems, the meta-metamodel and a set of algorithms made it possible to create a platform for building data consolidation systems [9]. The use of a software platform for building data consolidation systems ensures the systematization of the consolidated data and their consistent storage. In addition, the control model formed during the construction and development of the system is, in fact, a description of the subject area and can be reused.

One of the problems that have not been solved within the framework of the platform for building data consolidation systems is the automation of the creation of analytical reports. The existing solution is not optimal, the user defines the composition of the report and, together with the administrator, builds a view in the database. The metadata of the created view is saved using the tools built into the database and then read by the analytical module. The user is given an opportunity to view the data included in the report in the form of tables, graphs, and cartograms. The existing solution makes it difficult for users to independently manage analytical experiments, since it is associated with an appeal to the administrator or skills in working with SQL queries.

Using the control model allows us to create a tool that greatly facilitates the formation of analytical reports. The analytical reporting tool, as well as the entire model-driven system, should be based on specific models. The CWM specification [10] (Common Warehouse Metamodel) is the most widely used in the field of analytical data processing. The analytical model of the CWM specification is based on the fact-dimension model proposed by Mateo Golfarelli (et al.) in 1998 [11]. According to the proposed model, data are divided into aspects of analysis - dimensions, and aggregated numerical characteristics - measurements. In turn, the measurements are combined into facts, called cubes in the CWM terminology. An

algorithm for constructing an analytical model (dimensions, cubes, and connections between them) based on the data of the control model was developed and presented in [4].

An urgent task is to create a visual designer of exploration queries to the accumulated data using the resulting model. The following describes the algorithm for forming a user request to the database based on the analytical model and shows the interface for its formation.

2 Exploration Query Builder

2.1 Algorithm for Generating Exploration Queries

The existing analytical module of the platform for building data consolidation systems has been tested on several large tasks, about a hundred reports have been created in it. It is very costly to redo the existing reports, the same concerns the implemented primary analysis tools. Therefore, it is necessary to give users a tool which allows them to create reports in an intuitive manner in the format accepted in the system. But, besides this, it is necessary to save the description of the report (analytical model) in the form of metadata. The analytical model will allow us to edit and supplement reports in the future.

User's behavior when creating a report without using the constructor is described above, and it consists of several simple steps. First, the user defines the classes and attributes involved in building a report and the aggregation function, and then the administrator writes a request. In the exploration query builder, we need to facilitate the process of selecting classes and attributes, as well as to automate the construction of a view in the database.

In terms of the CWM specification, the exploration request generation algorithm is described in Fig. 1. According to this algorithm, at the first stage, the system generates a connected graph of cubes and displays it to the user. Showing only the cubes to the user can significantly reduce the amount of the displayed information. In the process of selecting the cubes, unrelated branches are hidden, and this eliminates the possibility of building an incorrect query.

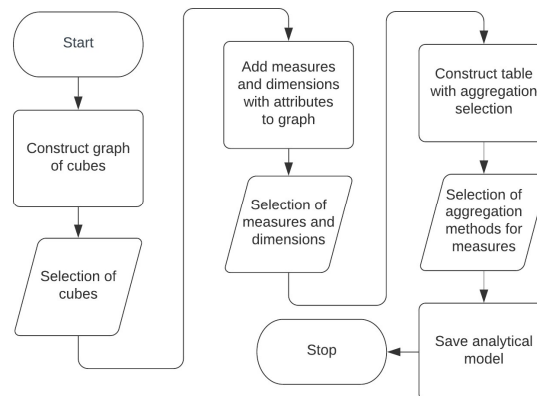


Fig. 1. Algorithm for generating a report by the user.

In the next step, the user sees the measures and dimension attributes of the selected cubes. The user, using the graphical interface, selects the aspects of analysis which interest him in the graph and defines the aggregation functions for the selected measures. Further, according to the choice, an expanded table with data is built. The user can assess the completeness of the sample and its compliance with the research objectives. The result of the work is a customized analytical model. The system saves the model to the database and automatically builds a view based on it. The view is made in accordance with the metadata accepted for the analytical module, and this allows one to see the resulting report in this module immediately after the creation or modification.

2.2 Exploration Query Builder Interface

The created exploration query constructor works according to the above algorithm. Let us consider its work using the example of the Soil Condition Research System. Initially, the system included the results of field studies, namely the assessment of the ecological state of the soil of the settlements of the Krasnoyarsk Region. At each of the objects of research full-profile soil sections were made, binding and description of the soils based on the morphological features were performed, the full name of the soil was given, and the soil samples were selected at different depths. To store this information, the following classes were created: "Sampling Place", "Soil section", "Soil type", "Section horizon", etc. The next step was to study the chemical composition of the natural samples, content of mobile nitrogen and phosphorus, and humus state. To store the information, classes of the same name were created.

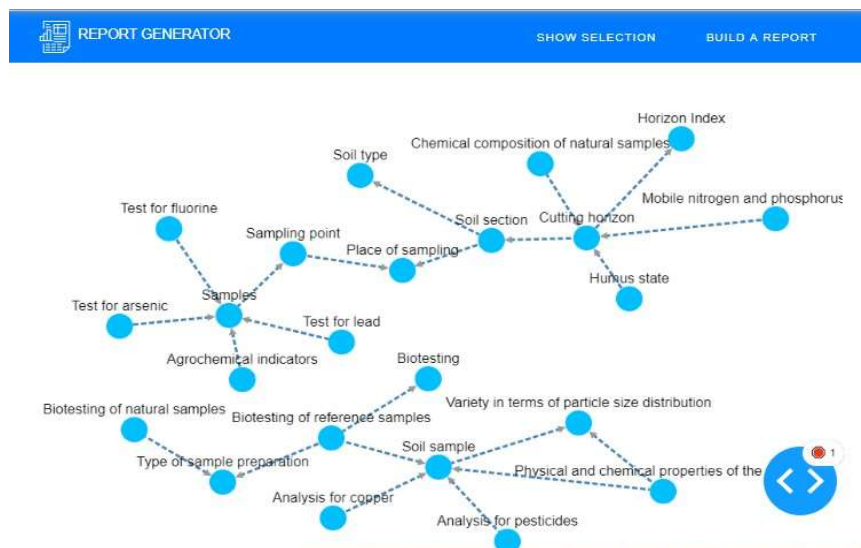


Fig. 2. Graph of the cubes.

Later, to study the dependence of the results of biotesting on the storage duration of the soil samples and conditions of the sample preparation, carried out by a group of students, a separate branch was created. The branch contained the classes “Biotesting of natural samples”, “Biotesting of reference samples”, etc. Unfortunately, the unsuccessful numbering of the natural samples did not allow them to be unambiguously compared with the data already available. The comparison is still underway.

To continue the study of natural samples, sampling and analysis of samples of technogenic surface formations near the industrial enterprises of the city of Krasnoyarsk were carried out. The samples were analyzed for the concentration of lead, arsenic and fluorine. Agrochemical indicators were studied. The control model of the system was supplemented with the appropriate classes.

Last year, within the framework of a grant the Krasnoyarsk reference center conducted an experiment with the reference soils to calibrate and determine the boundaries of the effectiveness of bioluminescent enzymatic analysis as a biotesting method to assess the degree of soil degradation, as well as anthropogenic and technogenic pressure on soils. The studies were carried out using various enzyme systems using standard methods of reference samples for common pollutants. The following classes were added to the system: “Analysis for pesticides”, “Analysis for copper”, “Physicochemical properties of the sample”, etc.

Thus, at present, the Soil Condition Research System contains data from five different experiments. An analytical model containing the information about all the cubes created during the development of the system can be seen in Fig. 2.

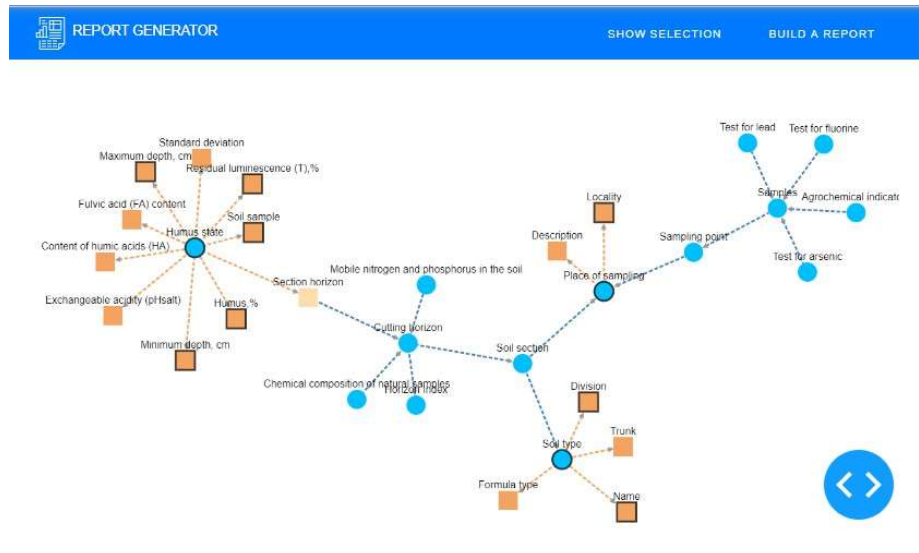


Fig. 3. Graph with the measures and dimensions.

At the first stage of constructing a research query, the user can select any cube from the proposed ones, but after selecting the first cube, the unrelated branch will disappear from the graph. For example, if you select the cube “Sampling location”, only the “Field studies” branch associated with this cube is available (Fig. 3). In Figure 3, we can also see that the user has selected the cubes: ‘Humus state’ and “Soil type”. The cubes selected by the user display the related elements; these are the dimension attributes and measures. With the help of simple mouse manipulations, the user can select attributes and indicators of interest for him. The cube – “Sampling location” has two attributes: “Settlement” and “Description”, the “Settlement” attribute is selected. “Soil Type” has four attributes. Have been selected: “Department” and “Name”. In the “Humus state” cube, four indicators. Have been selected: “Humus, %”, “Maximum depth, cm”, “Minimum depth, cm”, “Residual luminescence (T), %” and “Soil sample”.

After selecting the interesting aspects of the analysis, the user can view a detailed description of the selected attributes and indicators, start generating a database query based on them and examine the data obtained in detail (Fig. 4). In Figure 4, we can see that almost all the indicators are filled, but some of the data is not indicated. For example, the section of the soil type is empty, the locality is filled with service information. If necessary, one can add or edit these data in the data entry module. If the user is not satisfied with the composition of the report, he can return to the previous stage of selecting attributes and measures. If the data is selected correctly, the user proceeds to the selection of the aggregation function for the indicators, in our case it is “Humus,%”, “Maximum depth, cm”, “Minimum depth, cm”, “Residual luminescence (T), %” and “Soil sample”. After selecting the aggregation functions, the user starts the save process. As a result, the analytical model is saved in the

database and a report is created. The report is available for viewing in the analysis module.

REPORT GENERATOR		SHOW SELECTION		BUILD A REPORT			
Locality	Name	Section	Gumus,%	Minimum depth, cm	Maximum depth, cm	Residual luminescence (T),%	Soil sample
Minino	Chernozems	Accumulative-humus					
Unknown 1 (lead)	Chernozems	Accumulative-humus	5.99	12	26	37.98	8
Unknown 1 (lead)	Chernozems	Accumulative-humus	2.37	35	40	74.95	nine
Unknown 1 (lead)	Chernozems	Accumulative-humus	0.77	70	80	85.47	ten
Unknown 1 (lead)	Chernozems	Accumulative-humus	0.77	115	125	99.68	eleven
Unknown 1 (lead)	Leached chernozems		4.92	27	43	35.36	3
Unknown 1 (lead)	Leached chernozems		2.47	43	55	45.49	
Unknown 1	Leached						

Fig. 4. Table with the data.

The created custom exploration query builder works according to the proposed algorithm. It allows users to quickly conduct analytical experiments by creating new or editing old analytical models.

3 Conclusion

The platform for building data consolidation systems is used in many subject areas to support the collection, storage, and analytical processing of data. The proposed algorithm for the formation of exploration queries and the query designer built on its basis will facilitate the process of data analysis by the users and attract new researchers to use the platform.

Acknowledgments. The study was carried out with the financial support of RFBR and the Government of Krasnoyarsk region, research project №18-47-240005.

References

1. Korobko, A., Nicheporchuk, V., Nozhenkov, A.: Dynamic Generating User Interface of the Data Consolidation Web-system for Emergency Monitoring. *Informatization and communication* **3**, 59–64 (2014) (in Russian)
2. Korobko, A., Korobko, A.: Applied model of the scientific activity accounting system. In: *Proceedings of the International Scientific and Practical Conference on the Actual*

- Problems of Mathematical Modeling and Information Technologies, pp. 71–75, Sochi (2015) (in Russian)
3. Korobko, A., Korobko, A., Karepova, E. Certificate of state registration of a computer program No. 2018614158 dated 02.22.2018. Copyright holder: FRC KSC SB RAS
 4. Korobko, A., Korobko, A.: Constructing the analytical model for specialized model-driven system of scientific data consolidation. In: CEUR Workshop Proceedings, vol. 2534, pp. 377–383 (2019)
 5. Korobko A., Korobko, A.: Information Modeling of Temporal Spatial Data for Ecological Monitoring of the Krasnoyarsk Reservoir. In: CEUR Workshop Proceedings, vol. 2033, pp. 319–323 (2017)
 6. Object Management Group (OMG): Model Driven Architecture (MDA). MDA Guide Revision 2.0 (ormsc/2014-06-01), June (2014)
 7. Seidewitz, E.: What models mean. *IEEE Softw.* **20(5)**, 26–32 (2003)
 8. Korobko, A., Korobko, A.: An original approach to the construction of a model-driven data consolidation system. *Informatization and Communication* **4**, 232–238 (2017) (in Russian)
 9. Korobko, A., Korobko, A.: A software platform for constructing model-driven systems for primary data consolidation. In: Proceedings of the VII International Conference “Knowledge-Ontology-Theories” (ZONT-2019), pp. 203-212, Novosibirsk (2019) (in Russian)
 10. Peyton, L.: Common Warehouse Metamodel. In: LIU L., ÖZSU M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA (2009). https://doi.org/10.1007/978-0-387-39940-9_900
 11. Golfarelli M., Maio, D., Rizzi, S.: The Dimensional Fact Model: A Conceptual Model for Data Warehouses. *Int. J. Cooperative Inf. Syst.* **7**, 215–247 (1998)