Formation of geospatial information databases in the system of digital spatial data infrastructures

Anatoliy A. Yamashkin¹, Stanislav A. Yamashkin¹*, and Milan M. Radovanovic²,³

¹ National Research Mordovia State University, Bolshevistskaya Str., 68, 430005 Saransk, Russia
² Geographical Institute "Jovan Cvijic", Serbian Academy of Sciences and Arts, Djure Jakšića, 9, 11000 Belgrade, Serbia
³ South Ural State University, Lenin prospekt, 76, 454080 Chelyabinsk, Russia

Abstract. The article discusses the key aspects of the development of spatial data infrastructures and models of spatio-temporal data based on the study of geosystems, as well as ensuring the updating of the geospatial storage of information based on Earth remote sensing data. The authors have shown that for the formation of the software and hardware infrastructure of a digital storage of spatial data, it is advisable to follow an organized technological process. At the same time, spatial data warehouses should provide system integration of data with spatial and temporal reference. The solution to the problems of visualization and dissemination of spatial data should be based on the effective use of geoportal systems. It is concluded that the development of databases of digital SDIs and geoportal systems is associated with the development of methods and algorithms for the conjugate analysis of the peculiarities of the interaction of natural, social and production systems, complex interpretation of large arrays of spatial data and forecasting the development of natural and natural-man-made processes.

1 Introduction

The strategic goals of geosciences should include the development of methodological and algorithmic support for the functioning of spatial data infrastructures (SDI), based on the use of geosystem models of the structure and dynamics of development of natural-social-production systems (NSPS) and their interaction, assessment and forecasting of ecological-social-economic processes, natural and man-made emergencies in the geographical envelope [1]. The modern paradigm of earth sciences, responding to the challenges of aggravating global and regional geoecological problems, is focused on solving the following key tasks:

− development of end-to-end information technologies for integration, processing and analysis of spatial data, means of production of new digital cartographic models;
− increasing the efficiency of methods and algorithms for organizing databases of large arrays of spatially distributed data using multi-model DBMS;

* Corresponding author: yamashkinsa@mail.ru

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
development of strategies for the effective development of information resources to support management decisions in the field of sustainable development;

development of methodology and methods for analytical and synthetic mapping.

An important role in the analysis of geosystems in the structure of problem-oriented SDIs is assigned to spatial data, on the basis of which digital geographic maps are constructed. Their purposeful use in combination with Earth remote sensing (ERS) data and other information resources serves as a reliable source for building various types of models aimed at quickly and efficiently solving problems in assessing and predicting the ecological-socio-economic development of regions, making managerial decisions on optimization nature management, minimization of natural and technogenic emergencies.

The set of elements of the main cartographic and information layers forms the basis that provides the linking and processing of all thematic information functioning in the SDI. In general, all general geographical and thematic maps, remote sensing data constitute a single information complex capable of solving problems for which the problem-oriented SDI is intended. Each of these elements can influence the adoption of management decisions, justify, supplement them. At the same time, the base layers are involved in the development of almost all tasks at different stages of the modeling and solution preparation processes.

2 Materials and methods

The formation of a system of digital maps involves the development of methods and algorithms for the integration, analysis and practice-oriented use of cartographic information. The basic cartographic basis of digital SDI is defined as a certain set of mutually consistent maps and geo-images that are subject to mandatory and long-term storage in a spatial database and are used as a basis for modeling natural, social and production systems and their interactions [2]. The appeal to certain basic elements is determined by the range of problems to be solved to optimize the functioning of geosystems. The process of developing basic mapping software for SDI is based on the following algorithm:

- determination of the composition of elements of basic cartographic support;
- development of a system for collecting, entering, placing, storing, processing and visualizing basic cartographic information;
- creation of a multipurpose cartographic base to ensure the work of the SDI.

It is advisable to include in the content of the cartographic base of the territory common geographic objects characteristic of all groups of thematic maps, and supplement them with special elements that emphasize the characteristic features of the spatio-temporal organization of geosystems. In addition, it seems necessary to include in the basis the elements of spatio-temporal referencing, which provide opportunities for the interconnected input, storage, visualization and use of all thematic SDI information [3]. The main sources of updating the basic cartographic framework are remote sensing data, state statistics, and environmental monitoring. Current trends in the use of spatial data are focused on mathematical modeling and machine analysis, which involves the organization of navigation systems between digital models of natural, social and production systems.

The study of natural, social, production systems is based on the observance of genetic, historical and structural principles. Their observance presupposes diagnostics of the conditions for the emergence of the system under study, analysis of development stages, manifestations of periodic, cyclical, rhythmic processes and phenomena, assessment of the current state, as well as possible development prospects. This approach optimizes the clarification of internal connections and dependencies, interactions and relationships between the elements of systems, the disclosure of the internal organization of geosystems; regularities of the spatial organization of the territory, features of the development of natural, social, economic and ecological processes.
The most important geographic process in the geographic envelope is economic development - the saturation of ecosystems with various technogenic objects, as a result of which a certain type of interaction of natural, social and production systems is formed. As a result of the interaction of technogenic and natural processes, a specific structure of nature management is formed. At the same time, the negative consequences of the development of accompanying geocological processes are noted outside the zone of the actual location of anthropogenic systems. Due to this, the features of socio-economic development are deployed on the technogenic modifications of the geosystems of past periods.

The production of digital models in modern conditions is a multi-stage, relatively long process, in which cartographic information is converted from graphic form to digital, recorded on a machine-readable medium, and subjected to repeated processing and structuring [4]. Obviously, the main indicator of the quality of a digital model is its reliability - the degree of accuracy (without distortion) of reception, transmission, processing and storage of cartographic information in the system under the given conditions of its operation.

In modeling the spatio-temporal states of geosystems, the priority task is to study connections. In the geographic envelope, in addition to direct, reverse and indirect connections between natural, social and production systems, relations-relations and connections-connections are distinguished. Contact and remote connections between nature, population and economy in a specific area of space lead to changes in their states in time, which is conceptualized in the cultural landscape in the form of natural, cultural heritage, the existing system of land use, geocological problems, etc.

Detection of complex nonlinear relationships in geosystems and the determination of patterns from development is possible through the application of deep machine learning technologies to the analysis of spatio-temporal data and is used to solve project-oriented problems: monitoring the state of ecosystems [5], predicting the development of emergency processes and phenomena [6], analysis of spatially distributed social processes [7, 8]. The use of machine learning methods and algorithms makes it possible to increase efficiency and reduce research costs due to accurate extrapolation of measurements. In this direction, the use of ensembles based on the combination of deep architectures of various types, including convolutional and recurrent models, is relevant.

The development of models should be guided by the disclosure of the spatio-temporal organization of the hierarchical organization of geosystems of the geographic shell - the contact zone and active energy and material exchange between the lithosphere, atmosphere, hydrosphere and biosphere. This sphere is distinguished by the most active development of geographical processes: weathering, development of exogeodynamic, hydrological, soil-forming, biological processes; characterized by the highest concentration of life on Earth, the emergence and evolution of mankind.

Under the influence of the uneven distribution of solar energy, energy of intraterrestrial origin and the effects of self-development, the geographic envelope is subdivided into geosystems, which are distinguished by the relative homogeneity of the structure, mode of functioning, dynamics and development. Geosystems that are part of the structure of the geographic envelope have a hierarchical spatio-temporal organization. Depending on the goals of research and regional coverage, the systematization of information is carried out at the level of systems, classes, groups, types, genera and types of geosystems.

3 Results and discussion

As a result of preliminary experiments and initial generalization of the materials, we assumed that in the SDI, to assess the functioning of geosystems, it is advisable to distinguish the following groups of models:
Natural systems – structure, dynamics and functioning of the geographical envelope and its structural parts; electronic models characterize the energy and information properties of geospheres: geology, tectonics, hydrogeology, endogenous and exogeodynamic processes of relief formation; climate, surface waters and hydroclimatogenic processes; soil cover structure; biota and distribution of rare, endangered plant and animal species.

Social systems – the spatio-temporal structure of society, or society: population, demography, migration, ethnic groups, cities, countryside, settlement, recreation, social composition of the population, labor market, confessional composition, lifestyle, needs; problems and events of public life related to the functioning of states: electorate, borders, conflicts, countries; spatio-temporal structure of sustainable forms of human activity: cultural landscapes, humanitarian processes, culture, institutional environment - political, social and legal rules of production, exchange and distribution, spatial models of immanent philosophy.

Economic systems – the territorial organization of the company's economic activities: natural resources, industry, transport, agriculture, tertiary sector, innovation, investment complex.

Geoecological systems – spatio-temporal systems reflecting the structure of the geographic envelope as a habitat for humans and other organisms, their change in the conditions of technogenesis and measures to ensure sustainable development.

The most important areas of analysis are the complex characteristics of countries and regions, which systematize and summarize heterogeneous data on nature, population, economy, culture and social organization for various types of activities (educational, scientific, reference, military, etc.) and geographical forecasting – development notions of geographic objects inaccessible to direct research, primarily the spatial-temporal forecast of ecological-socio-economic development for making managerial decisions.

Effective land diagnostics based on the analysis of a complex hierarchy of geosystems is possible on the basis of thematic interpretation of spatio-temporal data through machine learning algorithms, modeling of system connections in natural-social-production systems using models; based on the hybrid use of recurrent and convolutional neural network layers, combined into deep structures; integration of large arrays of spatio-temporal data and knowledge based on cloud technologies.

It is advisable to form a block for analyzing the interaction of natural, social and industrial systems of project-oriented SDI on the basis of digital models of geofields. This form of spatial data presentation is most convenient for the restoration of the indicator at any point of the digital model, its cartographic display, mathematical statistical processing and spatial modeling. The development of convergence processes of GIS technologies and technologies for intelligent analysis of remote sensing data makes it possible to form the basis for spatial analysis and modeling of geosystems based on modern digital methods of processing remote sensing data. The indicated technologies are the most promising, since they help expand the capabilities of problem-oriented GIS as systems for processing and analyzing spatial-temporal information. The efficiency of analysis increases significantly when using databases that store large arrays of spatially distributed information, integrated from different sources.

One of the directions of development of geoinformatics in this case is the expansion of the functionality of SDI and GIS, as a key basis for the organization of modern scientific approaches to the assessment and analysis of the geographical envelope, providing broader prospects for using SDI with the support of managerial decision-making to implement the strategy of sustainable development of regions. At the same time, it is necessary to resort to a certain number of generalizations and abstractions, allowing to reduce all the richness and diversity of the surrounding space to the creation of geoinformation models, characterized by a finite number of natural objects and properties describing them.

To solve the problem of creating a software and hardware infrastructure for a digital storage of spatial data, it is advisable to follow an organized technological process, which...
begins with analysis and access to the strategic goals of deploying digital SDI, optimization of subsystems at various logical levels: equipment, data storage systems with various options for organizing access, virtual and physical servers, as well as system and application software. Spatial data warehouses should provide system integration of data with spatial and temporal reference. Among spatial DBMS, PostgreSQL with the PostGIS, MySQL Spatial, Oracle Spatial extensions should be distinguished. Having a spatial DBMS that supports a sufficient number of geometry types and spatial operators provides advantages for a spatial data center. The tasks of indexing and searching for spatial data, their system analysis and visualization within the SDI are interrelated [9]. This fact imposes restrictions on the structure of the data warehouse infrastructure. The architecture of the spatial database should be built on the basis of the set design tasks, solved by means of SDI [10].

The solution to the problems of visualization and distribution of spatial data integrated into digital SDI should be based on the implementation and effective use of architectural solutions in the field of construction and deployment of geoportal systems.

4 Conclusions

The analysis of problem points in the field of formation of geospatial information databases in the system of digital infrastructures of spatial data made it possible to draw the following key conclusions.

1) For the formation of the software and hardware infrastructure of a digital storage of spatial data, it is advisable to follow an organized technological process. At the same time, spatial data warehouses should provide system integration of data with spatial and temporal reference. The solution to the problems of visualization and dissemination of spatial data should be based on the effective use of geoportal systems.

2) The development of databases of digital SDI and geoportal systems is associated with the development of methods and algorithms for the conjugate analysis of the peculiarities of the interaction of natural, social and production systems, complex interpretation of large arrays of spatial data and forecasting the development of natural and natural-man-made processes.

3) Effective land diagnostics should be based on the analysis of a complex hierarchy of geosystems, which relies on thematic interpretation of spatio-temporal data through deep machine learning algorithms and modeling of system connections in the NSPS. It is advisable to form the NSPS analysis block for project-oriented SDIs based on digital models of geofields.

5 Acknowledgements

The reported study was funded by RFBR, project number 20-37-70055.

References

1. S.M. Vdovin, S.A. Fedosin, A.A. Yamashkin, S.A. Yamashkin, Natural hazards: the connection between science and practice: materials of the II Intern. scientific-practical conf, 82 (2015)
2. S.A. Yamashkin, Scientific and technical bulletin of Volga region, 6, 223 (2015)
3. M.F. Goodchild, International journal of spatial data infrastructures research, 2(2), 24 (2007)
4. I. Masser, *SDI Convergence: Research, Emerging Trends, and Critical Assessment*. *Netherlands Geodetic Commission* (2009)

5. W. Cheng, Y. Shen, Y. Zhu, L. Huang, Thirty-Second AAAI Conference on Artificial Intelligence (2018)

6. Y. Tao, X. Gao, A. Ihler, K. Hsu, S. Sorooshian, 2016 IEEE Congress on Evolutionary Computation (CEC), 1349 (2016)

7. Y. Li, K. Fu, Z. Wang, C. Shahabi, J. Ye, Y. Liu, Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 1695 (2018)

8. D. Wang, J. Zhang, W. Cao, J. Li, Y. Zheng, Thirty-Second AAAI Conference on Artificial Intelligence (2018)

9. L. Candela, D. Castelli, P. Pagan, ERCIM News, *(89)*, 37 (2012)

10. J. Lee, M. Kang, Big Data Research, *(2)*, 74 (2017)