Integration of building information modeling and big data processing technologies

Alexei Sirotskiy*

Moscow State University of Civil Engineering, 26, Yaroslavskoye shosse, Moscow, 129337, Russia

Abstract. The article deals with the issues of global integration of the accumulated arrays of information models of capital construction objects into a common space of information models, created according to the principles of organization and big data processing. The problems of geoinformation binding of information models, organization of means and places of their storage, methods of processing and providing access to interested subjects are considered. It is proposed to create a common space of information models according to a distributed decentralized scheme with direct interaction between state and commercial participants, both in terms of business integration, and in terms of joint data management and implementation of replication technologies.

Keywords: construction site, information modeling, big data, analytics, digitalization, systems approach, integration, model space

1 Introduction

Active digitalization and the introduction of information technologies in the fields of construction, the integrated development of cities and territories give rise to new tasks for the interconnected integration of digital technological solutions [1], which are due to the accumulation of huge amounts of data and the need for their analysis.

At the moment, great attention is paid to the development of information modeling technologies in construction (Building Information Modeling - BIM), which are mainly considered in the context of capital construction objects life cycle management, which will allow receiving and processing any necessary information about the object of interest at all stages of the life cycle and, importantly, significantly improve the reliability of the data and information used.

Unfortunately, while the widespread use of information modeling technologies in relation to construction projects, starting from the stages of their conception and design, is not observed. So far, separate measures are being taken regarding the introduction of BIM in the design and operation of facilities with state participation. Requirements for the use of information modeling in the design and construction of private construction projects have

* Corresponding author: hotwater2009@yandex.ru

not yet been developed at all. There are a number of objective reasons for this, in particular: the unavailability of the regulatory framework, the lack of standardization, the unavailability of the infrastructure of all participants and interested parties, the incomplete elaboration of technical issues about the mechanisms and features of the implementation of information modeling [2]. However, the systematic process of introducing BIM is steadily going on, and the transition of all processes of documenting, designing and managing real estate objects based on their digital models is only a matter of time.

Another driving process in the modern world is the rapid development of technologies for collecting, managing and processing big data (Big Data - BD). To date, database analytics has shown high efficiency in making management decisions in such industries as advertising, trade, logistics, industrial production, and many others. It is possible to predict great prospects for the analysis of databases accumulated in the entire set of information models of construction objects. Since information models can contain data of various kinds, both about the geometric parameters of objects, the properties of the materials used, and about the operation of objects, an analysis of the entire accumulated array of databases on real estate objects will reveal new knowledge and new patterns that can be used to create new objects, both in terms of goal-setting and the area of the most effective use, and in terms of the design ratio and construction costs in conjunction with the operational properties and objects parameters.

Optimization and improvement of construction processes based on the introduction of information modeling is widely discussed in a number of scientific papers. Thus, in [3], the problem of optimizing the construction time at the early stages of the life cycle of construction objects is considered by information modeling, coupled with the involvement of the customer himself in the active phase of construction, at least for planning work schedules. The digitalization of the construction industry with the use of BIM is considered more widely in the work of T.Yu. Ovsyannikova and A.A. Patsukov [4]. Here, the issue of combining various information models within a common digital platform and the transition to the creation of information models of cities (City Information Modeling - CIM) is already raised, which, according to the authors, will contribute to the formation of a system for the most efficient use of land and property resources in the interests of the state, population and business.

Of course, increasing requests for information modeling as a technological platform automatically raises a number of questions about the software and technical tools based on which projects can be implemented, and this tool should be as reliable and accessible as possible, with a minimum risk of losing the possibility of its long-term use. This problem is well considered in the work of D.V. Sbrodov and N.A. Ivanov [5], who conducted a study of the development of the domestic software market for information modeling.

Thus, at the moment, in a variety of works, the solution of subject problems and questions on the use of information modeling and the achievement of local goals is considered. At the same time, it is obvious that there is an emerging need to integrate individual information models, to establish their relationship with territorial planning, the tasks of managing territories, the effective development of cities and regions in general, which for the foreseeable period makes relevant the creation of integrated systemic approaches to the organization of information construction models and their processing by methods and technologies of BD research.

The purpose of this work is to study methods and form approaches to creating, based on a systematic approach, integrated solutions for the means and methods of processing BD accumulated in disparate information models of capital construction objects. The implementation of this goal will allow creating analytical systems aimed at the development

of territories, the implementation of state programs and services, improving the quality of maintenance of buildings and structures, as well as formulating recommendations on the properties and parameters of the designed objects. In addition, the integration of construction information models into a common data processing environment with reference to spatial and geographic location will significantly improve the quality of cadaster and accounting of real estate, as well as documenting all existing communications.

Within the framework of this goal, tasks are set about:

- choice of methods and technologies for BD processing in construction;
- data segmentation of information models;
- approaches to solving problems of geoinformation binding of information models;
- ways and means of data storage of information models.

The designated goal and composition of tasks are relevant in the light of the development and scaling of information technologies, the creation of digital twins of objects in the surrounding world.

2 Methods

The creation of a systematic integrated approach to the accumulation, processing and analysis of data on capital construction facilities will significantly improve the efficiency and quality of managerial decisions on the infrastructure of cities and regions development, reasonably prioritize the number, purpose of the designed facilities and their quality and operational characteristics, find the optimal balance between the costs of designing and erecting objects and their ergonomic properties.

Research methods include:

- study, analysis and generalization of published scientific works in the field of the subject implementation of the implementation of BIM in construction;
- application of system analysis and a systematic approach to the study of scientific problems, structural and mathematical modeling of phenomena and processes, the theory and practice of digitalization of economic systems, including the digital transformation of enterprises, industries, regions, the introduction of end-to-end digital technologies in industry, finance and commerce, the implementation of the concept "Industry 4.0" [6];
- principles of digitalization of large data arrays;
- methods of multivariate BD analysis;
- means and methodology for processing BD [7] and large amounts of information, the theory and methodology of the object-oriented approach in the classification of research objects, methods for building databases and knowledge bases.

Of great importance for considering the integration of information models with BD processing systems and procedures is the standardization of the composition and presentation of data in information models, which is due to various existing standards in the field of information modeling, some of which are still under development [8].

An essential issue is the composition and extensibility of data that can be placed in information models and the definition of model types. So, at the moment, at least two lines of IFC formats versions 2.3 and 4 are simultaneously used. Each of them has its own subsets of model views. Moreover, in practical cases, it may be necessary to introduce new attributes into information models [9].

In general, methodological approaches to BD processing in construction, of course, have already been developed and discussed; they form the basis of this study:

- in the article [10], the principles of developing an information system for collecting and processing BD in construction are proposed;

- in the conference materials [11], solutions for using the BD for organizational and technological support when performing repair and construction work are proposed;
- article [12] is devoted to a review of BD integration methods in the framework of the digitalization processes of the construction industry.

It can be noted that the integrative processes of the database on construction objects are considered in relation to a variety of tasks. The methods applied by the authors mentioned above are focused on both local and global understanding of the BD.

In the local sense, the tasks of processing a BD of a corporate or other limited accessible segment and its corresponding data sets are considered.

In a global sense, there is a need to integrate different sets of data held by different entities.

The methods applicable to the analytics of the BD of construction objects, and allowing the use of artificial intelligence for decision-making, include:

- classification, clustering, segmentation;
- forecasting, building regression models;
- determination of the totality of the best characteristics of real estate objects;
- identification of associations between real estate objects and their environment;
- determination of the need for types and types of real estate objects;
- building relationships between information models of various objects based on new knowledge.

3 Results

The theory of information modeling of construction objects considers information models by type, in particular, architectural, design, calculation, engineering, estimate, production, operational, and also by dimension. The dimension of models in the most general case is divided into:

- two-dimensional models, which are actually digital analogues of classical design and drawing documentation;
- three-dimensional models that allow obtaining and building real images of object structures, including from any viewing angle;
- four-dimensional models that allow you to track the dynamics of the state of the simulation object in time;
- five-dimensional models that take into account the economic and cost indicators of the object;
- six-dimensional models, which reflect the actual state of the object at the current time;
- seven-dimensional models containing operational information about the object.

Within the framework of the type and dimension of the model, the volume and nature of the information that can be contained in this model becomes clear. The indicated principles of model typing well determine the composition of the data, but do not determine the spatial boundaries of the models. If we consider the information model in isolation, then this is not important. However, if we talk about the integration of information models into a common space of information models, then there is a need to determine at least the geometric boundaries of information models. On fig. 1 shows for example three adjacent buildings *A*, *B* and *C*, suppose that each of them has an information model.

A number of questions arise:

- about mutual orientation of information models;
- on the territorial binding of information models to the local plan;

-about interface of buildings information models with communications information models.

Thus, the information model of building *C* has design solutions for the adjacent territory, and with further mutual integration of this model, spatial and geometric overlays of this model on the territory of a neighboring object may occur.

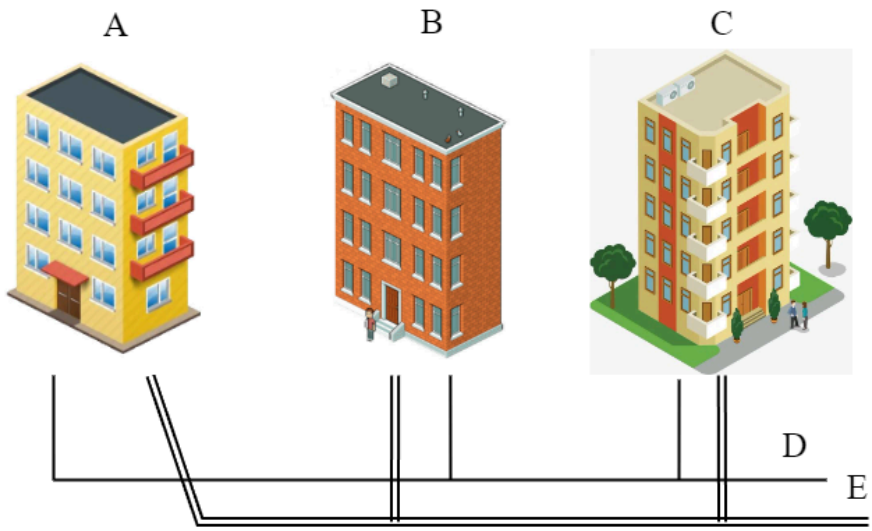

Fig. 1. Schematic representation of the adjacent information models integration.

When considering these models in isolation, this is not of interest; however, when solving problems of integrating models into a common model space, such overlaps are unacceptable, because they will represent geometric conflicts.

Similarly, the question arises about the interface of communications: what part of the engineering systems *D*, *E* will relate to the information model of the object, and what part - to the information model of citywide engineering systems and networks (the creation of information models of centralized systems and networks has not yet been carried out, but they will certainly will appear in the foreseeable future)?

In this case, it is necessary to determine both the methods for pairing model elements along their boundaries, and the methods for linking objects to a common model space.

In this case, a common space of information models (CSIM) is understood as a matrix geometric information base, to which information models of specific objects can be attached as components.

The matrix information base should describe the territory, the profile of the earth's surface and underground structures.

The biggest problem is the choice of reference coordinate system for the geometric matrix base. If, to create an information model of a detached building, it is quite acceptable to consider the territory conditionally flat and to carry out modeling in the Cartesian coordinate system, then for imposing the entire set of created models on the conditional surface of the earth and integrating them into the CSIM, this assumption will be unacceptable due to the high error at large distances and distortion of territorial plans. A clear and correct binding of information models to the territory is an extremely important aspect for the use of BD processing methods in cadastral engineering tasks.

To position an object on the earth's surface within the framework of the CSIM, you can use:

- system of geographical coordinates;
- development of the earth's surface on a Cartesian plane with non-linear scales of the coordinate axes.

Let us consider these options.

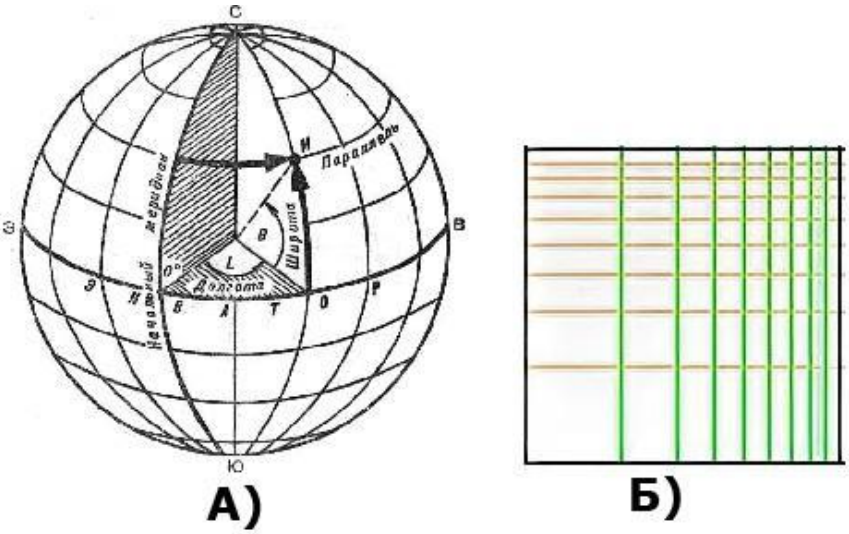

Fig. 2. Choice of a coordinate system of information models' common space.

The implementation of binding to geographic coordinates (fig. 2, A) looks quite reasonable, but it is fraught with difficulties: the fact is that in all information modeling systems, the construction of the geometric part of models of buildings and structures is carried out in the traditional Cartesian coordinate system. With the filling of the CSIM, eventually a conflict of accumulated errors will arise, since it will be impossible to make a close joint of two projection rectangular models on a convex surface without overlaps. Another difficulty arises due to the topography of the earth's surface, which is not yet taken into account here. To consider this fact, it will be necessary to introduce the concept of altitude, tied, for example, to the level of the world ocean. Although this does not present such a technical difficulty, it introduces a shift in the coordinates of the models, which in the modeling software are tied to the elevation of the earth's surface.

Another approach may be to choose a plane as the basis of the CSIM matrix (fig. 2, B). In this case, to eliminate discontinuities, it is necessary to switch to non-linear coordinate axes, where the scale of the axis depends on the coordinate. Modern software and hardware will make it quite easy to recalculate the coordinates, the conflict overlay of models will be practically excluded, but a high discreteness of the grid is required, and this representation is difficult to perceive for a person. However, the advantage can be noted that such a sweep can repeat the relief surface of the earth, and subject to this condition, there will be no need to shift the models in height.

In practice, it seems appropriate to create a common array of the model space according to the second option, but for the human-machine interface, it is still necessary to convert models to the first option.

The state or an authorized structure can act as the CSIM operator, common integrator and owner of the geospatial matrix. However, storing all model information with a common

operator, even at distributed data centers, is hardly possible. Considering that, information models can contain, among other things, a lot of additional graphic information, as well as replenished with operational data, the set of models can be considered as a digital twin of geospatial and construction data covering the entire territory in the future, and this will be expressed in such volumes of information that up to have not yet been processed in any system. In this regard, a common system can be considered for storing model bindings, model references, as well as a minimum amount of key design and control information. The CSIM itself should be built according to decentralized principles.

In the context of considering the issue of creating a CSIM, it is necessary to create successive links with the common data environment (CDE - Common Data Environment) of existing information models, since CDE is considered as the basis for information support for managing the life cycle of capital construction projects [13]. CDE is recommended for use in order to ensure operational access to information model data for all interested participants, their consistency, integrity, consistency, relevance, reliability and long-term storage. CDE is developed at the level of electronic document management and can be organized using various information systems, network local and external resources, cloud solutions, open data portals.

Thus, the creation of an CSIM as a database array, in particular, includes the main task of integrating various CDE with the definition of means of interaction, access and differentiation of user rights.

It is possible to propose the creation of the CSIM as a geographically distributed system consisting of interacting nodes. In this case, one common coordination center can be created, with which the nodes of the system, which are data storages, will interact. The proposed organizational and hierarchical system of CSIM nodes is shown in fig. 3.

By affiliation, nodes can be:

- departmental subordination, implemented on the basis of information systems of urban and municipal significance, and storing information about objects created as part of the implementation of programs with state participation;
- commercial, implemented on the basis of data centers of design construction organizations, and storing information about all other construction projects.

According to the territorial principle, nodes can be:

- federal level,
- regional level,
- urban level,
- municipal level.

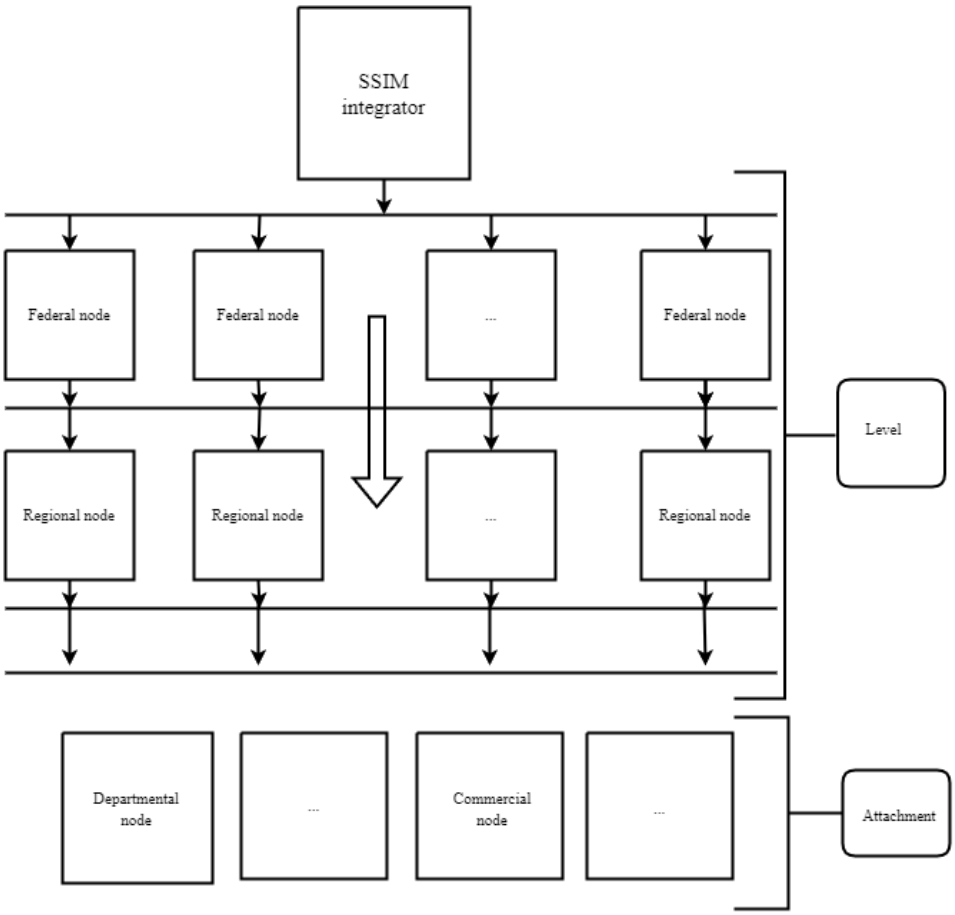

Fig. 3. The proposed structure of a common space of information models organization.

It is advisable to organize the principle of access to integrated data through a single entry point with access primarily to the CSIM portal, which contains an associated grid of objects and models and pointers to them. The pointers are redirected respectively to a departmental or corporate data server, from which an authorized user can extract the necessary part of the data according to the allowed definitions of the model view (MVD - Model View Definition).

Large data arrays on capital construction projects and related infrastructure through the CSIM will solve the following tasks (but not limited to):

- cadastral registration;
- maintenance of real estate register;
- maintenance of the real estate owners register;
- technical inventory of real estate;
- study of the need for new real estate;
- designing new real estate objects [14];
- coordination of building reconstruction [15];
- keeping a register of networks and communications;
- detailing cartography;
- analytics of passenger traffic and traffic congestion.

Within the framework of these tasks, it is necessary to ensure the possibility of one-time access to BD arrays to a huge number of users: both government agencies, design and service organizations, and citizens. This, in turn, requires high-bandwidth data access channels, which requires duplication and distribution of BD arrays to their storage locations. It is proposed to use the "master-slave" replication mechanism (fig. 4), which differs in that each node is the "master" of the set of information models assigned to it, and at the same time, it is a secondary node for one or more other data servers. Moreover, the number of secondary nodes to each main data server ("master") can be in the tens.

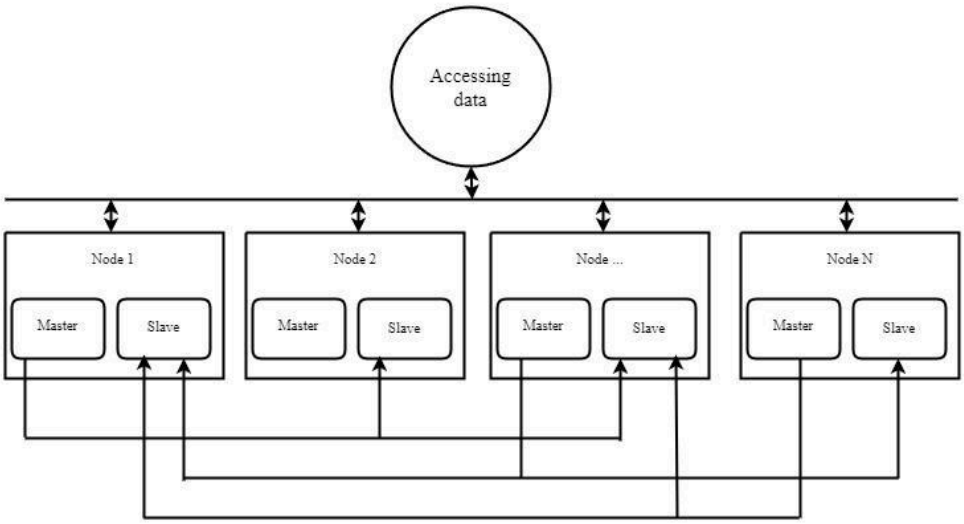

Fig. 4. The proposed mechanism for replicating big data of information models of capital construction projects.

The last of the issues raised in the framework of the study is the segmentation of data by access levels. Information models include:

- data on the structures and spatial elements of the building;
-information about materials, their grades and physical parameters;
- geometry and shaping of elements of building structures;
- parameters of premises as objects of operation;
- information about the dynamics of the facility, ongoing repairs and other activities.

For example, there can be hundreds of generalized operational and ergonomic parameters that can characterize a real estate object and which may be of interest for its comparison with other similar objects [16]. In total, for each real estate object there can be millions of features (attributes) that characterize it. These signs can be in the form:

- numerical parameters;
- text records;
- design documents in vector representation;
- photographic materials in the form of raster images;
- marked up machine-readable documents (for example, XML format);
- electronic non-broadcast documents (for example, in pdf format).

In addition, signs, and their corresponding entries, can be:

- fixed and unchangeable (for example, the year of construction);
- changeable (for example, the current deterioration of the building);

- replenished (for example, information about the work performed).

Some of them are of a purely technical or technological nature, which does not carry anything for the mass user of open data portals (for example, the brand of welding electrodes used to connect metal structures), and uploading such data to a common array is impractical, or even some of the information may be confidential information.

Based on this, each feature (group of features) should be assigned an access level and/or a presentation level. The access level determines to whom and in what cases this data can be freely provided. The presentation layer determines to which systems and/or organizations the information stored in a particular attribute is allowed to be transferred. Similar level values are assigned to users, organizations, and interacting systems. The model of such interaction is shown in fig. 5.

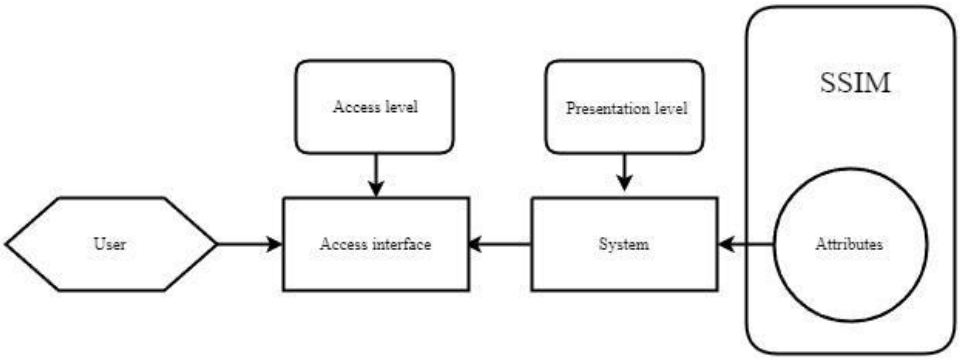

Fig. 5. Model of control and distribution of access to data.

The distribution model of levels and access rights will allow not only segmenting data based on their diversity, preventing overloading systems by providing information that is redundant for the user, but also partially solving the security problem of this multi-user system [17].

4 Discussion

Many processes of informatization and digitalization moved along the path of creating disparate information systems, services, and models, databases, with their subsequent unification, integration and creation of interaction channels. A typical example is the emergence of a government services portal, which has created a single entry point and a single access center to many services, including those related to property management. The system of electronic public services solves managerial and communication tasks and is centralized in its meaning. However, this example of integration shows that it is possible and necessary to follow leading trends in the creation of other systems of a similar scale. While the number of information models of construction objects is not so large, there is time for a detailed study of a complex of software and hardware solutions and the preparation of a regulatory framework for creating a common space of information models (CSIM). A distinctive feature of the CSIM will be its decentralized structure with many participants performing various roles and interacting electronically with each other. The CSIM will be able to act as a backbone for cadastral registration, cartography, planning and development of territories, a unified register of real estate, thus ensuring the provision of uniform and reliable information.

5 Conclusions

Within the framework of the study, achievements in the development of information modeling technologies in construction with big data processing technologies are summarized, tasks that can be solved by integrating these technologies are identified, and approaches are proposed to create a common space of information models as a decentralized distributed system, which in the future can become the basis reliable information about the entire set of information-construction and geospatial infrastructure and establish relationships between the tasks of cartography, accounting and registration of real estate, planning and development of territories.

References

1. Sirotskiy A.A., Samadurov A.E. Trends in the development of information services in the structure of the digital economy // Modern information technologies in education, science and industry: XI International Conference, IX International Competition of scientific and scientific-methodical works, Moscow, 02–03 November 2018 / Managing editors: T.V. Piryazeva, V.V. Serov. – Moscow: Sputnik+ Publishing House LLC, 2018. – p. 169-172. – EDN YWFBJB. (in Russian).
2. Talapov V.V. BIM technology. The essence and features of the implementation of building information modeling / - M.: Publishing house "DMK Press", 2015. - 410 p. (in Russian).
3. Oparina L.A. Application of information modelling technologies for construction time management // Smart Composite in Construction. – 2021. – Vol. 2, No. 2. – p. 48-55. – DOI 10.52957/27821919_2021_2_48. – EDN KNJPXR.
4. Ovsiannikova T. Yu., Patsukov A.A. Building information modelling systems: strategic objectives and realities of digital transformation in construction // Real Estate: Economics, Management. – 2022. – No. 1. – p. 13-18. – EDN VJSJCW.
5. Sbrodov D.V., Ivanov N.A. Software Support for Information Modeling in Russian Construction During the Sanctions Period // Components of Scientific and Technological Progress. – 2023. – No. 3(81). – p. 37-42. – EDN UDKXTZ.
6. Babkin A.V. etc. Digitalization of economic systems: theory and practice. – St. Petersburg: Polytech-Press. 2020. – 796 p. – ISBN 978-5-7422-6931-1. – DOI 10.18720/IEP/2020.3. (in russian).
7. Mayer-Schenberger V., Kukier K. Big data. A revolution that will change the way we live, work and think. Moscow: Mann, Ivanov and Ferber. 2014. - 240 p. ISBN 978-5-91657-936-9. (in Russian).
8. Ministry of Construction of Russia. Standards Development Plan. TC 10 / PC 8. Documentation for construction. URL: https://www.faufcc.ru/international-co-operation/international-organizations/plan-razrabotki-standartov/. (Retrieved 08/06/2023). (in Russian).
9. Ignatova E.V., Makisha E.V., Utkin M.A., Mochkin K.A. Creation of additional attributes of elements of information models based on the IFC format // Collection of materials of the All-Russian Scientific and Practical Conference “System Engineering of Construction. Cyber-Physical Building Systems - 2019”. -M.: NRU MGSU, 2019. - p. 217-222. (in Russian).
10. Istratova E., Sin D., Strokin K. Development of an Information System for the Collection and Processing of Big Data in Construction // Lecture Notes in Networks

and Systems. – 2021. – Vol. 149. – p. 189-196. – DOI 10.1007/978-981-15-7990-5_18. – EDN WTKYLP.

11. Ivanov N., Gnevanov M. The use of "big data" in the organization of repair and construction works to ensure OTR // E3S Web of Conferences: 24, Moscow, 22–24 апреля 2021 года. – Moscow, 2021. – DOI 10.1051/e3sconf/202126302010. – EDN SOARVT.
12. Yousif O.S., Zakaria R.B., Aminudin E., et al. Review of Big Data Integration in Construction Industry Digitalization // Frontiers in Built Environment. – 2021. – Vol. 7. – p. 770496. – DOI 10.3389/fbuil.2021.770496. – EDN CKSSAC.
13. Losev K. The common data environment features from the building life cycle perspective. IOP Conference Series: Materials Science and Engineering. 2020. Volume 913, № 042012, DOI:10.1088/1757-899X/913/4/042012
14. Kuzina O.N., Chulkov V.O. Information technology for the formation of an order for construction reorganization in an interactive mode // Industrial and civil construction. - 2011. - No. 3. - p. 41-42. – EDN NDXZWX. (in Russian).
15. Kuzina O.N. Coordination of the construction reorganization of non-production facilities // Innovations in the sectors of the national economy as a factor in solving the socio-economic problems of our time: Collection of reports and materials of the 2nd International Scientific and Practical Conference, Moscow, December 05–06, 2012 / Institute of Lifelong Education, Moscow State University Academy of Public Utilities and Construction, Central Research Institute of Economics and Management in Construction. Volume 1. - Moscow: Non-state educational institution of higher professional education "Institute of Continuing Education", 2012. - p. 201-204. – EDN TLUECD. (in Russian).
16. Sirotskiy A.A. Application of data analytics methods to assessing the prospects of real estate objects planned for construction // Construction: science and education. 2023. Vol. 13. Issue. 2. Art. URL: http://nso-journal.ru. DOI: 10.22227/2305-5502.2023.2. (in Russian).
17. Korneev A.A., Sirotskiy A.A. Security problems of the network infrastructure of multi-user design systems operating in construction // My professional startup: Collection of articles based on the materials of the X All-Russian Scientific and Practical Conference, Nizhny Novgorod, March 30, 2023. - Nizhny Novgorod: Federal State Budgetary Educational Institution of Higher Education "Nizhny Novgorod State Pedagogical University named after Kozma Minin", 2023. - p. 318-321. – EDN IUPQXV. (in Russian).