

УДК 004.9

ИНТЕГРАЦИЯ, ХРАНЕНИЕ И ОБРАБОТКА БОЛЬШИХ МАССИВОВ ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ ИНФОРМАЦИИ В ЦИФРОВЫХ ИНФРАСТРУКТУРАХ ПРОСТРАНСТВЕННЫХ ДАННЫХ

Ямашкин С.А., Ямашкин А.А.

*ФГБОУ ВО «Национальный исследовательский Мордовский государственный университет
им. Н.П. Огарёва», Саранск, e-mail: yamashkinsa@mail.ru*

Статья посвящена решению научной проблемы обеспечения процесса эффективной интеграции, хранения и обработки больших массивов пространственно-временной информации в цифровых инфраструктурах пространственных данных (ИПД). Авторами предложено в качестве ключевого аппарата в проектировании цифрового хранилища ИПД использовать геосистемный подход, показано, что для оценки процессов функционирования природно-социально-производственных систем целесообразно выделить три уровня интеграции и использования пространственно-временных данных: базовых пространственно-временных данных, пространственных моделей и принятия управленческих решений. К системным компонентам пространственных моделей отнесены модули, описывающие природные, социальные, экономические и геоэкологические системы. Уровень принятия решений предполагает проектно-ориентированное использование данных для организации процессов географического прогнозирования и формирования комплексной характеристики регионов. Авторами доказано, что для организации систем интеграции, хранения и обработки больших массивов пространственно-временной информации в цифровых ИПД целесообразно комплексное использование научно обоснованного набора систем управления базами данных, интегрированных в мультимодельное хранилище, дана характеристика доменам использования реляционных и NoSQL-хранилищ. Следствием использования мультимодельных хранилищ являются усиление качественных характеристик цифровой ИПД – гибкое масштабирование хранилища по мере необходимости, повышение надежности и отказоустойчивости.

Ключевые слова: пространственные данные, инфраструктура пространственных данных, система управления базами данных, мультимодельные хранилища, геоинформационные системы

INTEGRATION, STORAGE AND PROCESSING OF LARGE ARRAYS OF SPATIO-TEMPORAL INFORMATION IN DIGITAL SPATIAL DATA INFRASTRUCTURE

Yamashkin S.A., Yamashkin A.A.

National Research Ogarev Mordovia State University, Saransk, e-mail: yamashkinsa@mail.ru

The article is devoted to solving the scientific problem of ensuring the process of effective integration, storage and processing of large arrays of spatio-temporal information in digital spatial data infrastructures (SDI). The authors proposed to use the geosystem approach as a key apparatus in the design of the digital storage of SDI, it is shown that to assess the processes of functioning of natural-social-production systems, it is advisable to distinguish three levels of integration and use of spatio-temporal data: basic spatio-temporal data, spatial models and acceptance management decisions. The system components of spatial models include modules that describe natural, social, economic and geoecological systems. The decision-making level involves the project-oriented use of data to organize the processes of geographic forecasting and the formation of a complex characteristic of regions. The authors have proven that for the organization of systems for the integration, storage and processing of large arrays of spatio-temporal information in digital SDI, it is advisable to use a scientifically grounded set of database management systems integrated into a multi-model storage, and characterize the domains of using relational and NoSQL storages. The consequence of using multi-model storages is to enhance the quality characteristics of the digital SDI – flexibly scale the storage as needed, increase the reliability and fault tolerance.

Keywords: spatial data, spatial data infrastructure, database management system, multi-model storage, geographic information systems

Стратегическая цель развития инфраструктур пространственных данных (ИПД) заключается в формировании системного инструмента, позволяющего моделировать структуру и динамику природных, социальных, производственных систем (ПСПС), особенности их взаимодействия, проводить оценку эколого-социально-экономических процессов, прогнозировать развитие природно-техногенных чрезвычайных ситуаций в географической оболочке. В качестве ключевого понятия в проектиро-

вании ИПД целесообразно определить геосистему – «...особый класс управляющих систем; земное пространство всех размерностей, где отдельные компоненты природы находятся в системной связи друг с другом и как определенная целостность взаимодействуют с космической сферой и человеческим обществом» [1]. При более общем подходе геосистемы представляют собой тотальные системы, являющиеся «в равной мере географическими, экономическими, социальными и техническими», включаю-

щими многообразие экологических, социальных, экономических объектов, процессов, явлений и их взаимосвязи.

Важная роль в моделировании геосистем в структуре проблемно-ориентированных ИПД отводится географическим картам. Их использование в сочетании с данными дистанционного зондирования и другими информационными ресурсами служит надежным источником для построения различных типов моделей, ориентированных на эффективное решение определенного спектра задач по оценке и прогнозированию эколого-социально-экономического развития регионов, принятие управленческих решений по оптимизации природопользования, минимизации природно-техногенных чрезвычайных ситуаций.

Цель исследования заключается в решении задачи хранения информации о геосистемах с цифровых ИПД, сопряженной с необходимостью разработки и внедрения алгоритмов интеграции знаний, хранения и обработки больших массивов пространственно-временных данных. Важно подчеркнуть актуальность получения следующих конкретных результатов: 1) разработка эффективных физических и логических моделей хранилища больших пространственно-временных данных, основанных на мультимодельной организации систем управления данными; 2) создание новых алгоритмов кэширования и индексирования больших массивов данных для обеспечения высокой скорости их выборки и анализа; 3) внедрение архитектур программных интерфейсов для обеспечения необходимых выборок и группировок данных для последующего машинного анализа и визуализации.

Материалы и методы исследования

Базовое картографическое обеспечение проблемно-ориентированных ИПД целесообразно определить как системный набор взаимосогласованных карт и геоизображений, которые подлежат обязательному и долгосрочному хранению в пространственной базе данных и используются в качестве основы для моделирования природных, социальных и производственных систем и их взаимодействия [2]. Обращаемость к тем или иным базовым элементам определяется спектром решаемых проблем по оптимизации функционирования природно-социально-производственных систем.

Процесс разработки базового картографического обеспечения ИПД включает в себя следующий алгоритм: 1) определение элементов базовой картографической основы; 2) разработка системы сбора хранения,

обработки и использования базовой картографической информации; 3) создание картографической базы многоцелевого назначения для обеспечения работы ИПД: серий тематических карт природных, социальных и производственных систем; 4) определение последовательности и частоты доступа к базовым тематическим картам при решении определенных эколого-социально-экономических проблем.

В содержание картографической базы территории целесообразно включать общие географические объекты, характерные для всех групп тематических карт, и дополнять их специальными элементами, подчеркивающими характерные особенности пространственно-временной организации исследуемых объектов, процессов или явлений. Кроме того, представляется необходимым включить в основу элементы пространственно-временной привязки, которые обеспечивают возможности для взаимосвязанного ввода, хранения, представления и использования всей тематической информации ИПД. Очевидно, что все разнообразие сложных характеристик географических объектов, составляющих основное содержание проблемно-ориентированной ИПД, затрудняет их полное отображение на одном слое пространственной базы данных. В связи с этим наиболее целесообразным является расчленение этого комплекса на ряд специализированных базовых картографических слоев.

Совокупность элементов основных картографических и информационных слоев образует основу, которая обеспечивает связывание и обработку всей тематической информации, функционирующей в ИПД. В целом все общие географические и тематические карты, данные дистанционного зондирования составляют единый информационный комплекс, способный решать проблемы, для которых предназначена проблемно-ориентированная ИПД. Каждый из этих элементов может влиять на принятие управленческих решений. В то же время базовые слои участвуют в разработке практически всех задач на разных этапах процессов моделирования и подготовки решений. Основными источниками обновления базовой картографической основы являются данные дистанционного зондирования, государственной статистики, мониторинга окружающей среды.

Современные тенденции в использовании пространственных данных ориентированы на моделирование – организацию систем навигации между природными, социальными и производственными системами и их элементами. Общая схема раз-

вертывания работ по формированию ИПД включает: 1) разработку моделей в вербальной, графической, математической форме; 2) создание математико-картографических моделей; 3) сопоставительный анализ новых знаний, проверка моделей на практике.

В глобальном измерении объектом моделирования выступает географическая оболочка – зона контакта и активного взаимодействия литосферы, атмосферы, гидросферы и биосферы. В его составе выделяется ландшафтная оболочка – зона непосредственного взаимодействия геосфер, нижняя граница которой проводится по подошве первого подземного потока, а верхняя граница – по высоте влияния подстилающей поверхности на формирование местного климата.

В региональном аспекте объектами моделирования выступают геосистемы, которые обособляются в географической оболочке под влиянием неравномерного распределения солнечной энергии, энергии внутриземного происхождения и эффектов саморазвития. Они отличаются относительной однородностью структуры, режима функционирования, динамики и развития. Геосистемы имеют иерархическую пространственно-временную организацию. В зависимости от целей исследований и регионального охвата систематизация информации проводится на уровне систем, классов, групп, типов, родов и видов геосистем [2]:

– *системы*, выделяемые по особенностям макроклимата и определяющие особенности развития геоэкологических процессов: выветривание, рельефообразование, гидрологические, почвообразовательные процессы, биологический круговорот;

– *классы геосистем*, обусловленные тектоникой макроформы рельефа, определяющие перераспределение тепла и влаги, интенсивность экзогеодинамических процессов, а следовательно, проявление высотной поясности и вертикальной зональности;

– *группы геосистем*, определяющие функционирование литогидрогенных систем, развитие радиального и латерального водообмена, типа водно-геохимического режима;

– *типы геосистем*, формирующиеся на основе особенностей развития почвообразовательных и биологических процессов;

– *роды геосистем* – отличающиеся пространственными закономерностями развития экзогеодинамических процессов (эрозионных, карстовых, суффозионных, оползневых и т.д.) и морфоскульптурных форм рельефа.

Ландшафтная оболочка и иерархия геосистем – это арена развития глобальных, региональных и локальных процессов хозяйственного освоения. Насыщение геосистем различными техногенными объектами сопровождается формированием определенного типа взаимодействия природных, социальных и производственных систем. В этой связи приведем формулировку К.П. Космачева: «Общество всегда является активной стороной, определяющей тип развития территории.... Но результат развития во многом зависит от естественной основы территории, от того, как оно «воспринимает» влияние общества, насколько оно способно накапливать результаты человеческого труда и сохранять их надолго» [3]. В результате взаимодействия природных и техногенных процессов создается специфическая структура землепользования, сопровождающаяся активацией специфического спектра геоэкологических процессов. Негативные последствия развития могут быть отмечены за пределами зоны фактического размещения техногенных систем (загрязнение окружающей среды, истощение подземных вод). В силу этого особенности социально-экономического развития развертываются на техногенных модификациях геосистем прошлых периодов.

В моделировании пространственно-временных состояний геосистем приоритетной задачей является исследование связей. В географической оболочке кроме прямых, обратных и косвенных связей между природными, социальными и производственными системами выделяются связи-отношения и связи-взаимодействия, особенности которых в контексте физико-географических исследований раскрыты Н.А. Солнцевым [4]. В трактовке автора связь-отношение есть «всякая форма соотвествия явлений, которая может быть представлена в виде функциональной зависимости переменных величин, имеющей либо однозначный характер, либо вероятностный», а связь-взаимодействие – «это процесс взаимного воздействия объектов, протекающий в реальном пространстве и времени и выраженный в определенном ходе событий, которые происходят с веществом и энергией на определенном участке пространства». Контактные и дистанционные связи между природой, населением и хозяйством на конкретном участке пространства приводят к изменениям их состояний во времени, что концептуализируется в культурном ландшафте в виде природного, культурного наследия, сложившейся системы землепользования, геоэкологических проблем [5].

Диагностика и оценка связей в сложной иерархии геосистем является важнейшим направлением в разработке и практическом использовании новых автоматизированных методов и алгоритмов интерпретации данных ДЗЗ и вспомогательной пространственно-временной информации с использованием машинного обучения в рамках ИПД [6], постоянной актуализации облачных хранилищ больших массивов пространственно-временных данных, для организации которых актуально использование мультимодельного подхода [7]. Решение задачи эффективного распространения пространственных данных достигается посредством внедрения геопортальных систем [8, 9], характеризуется свойством проектной ориентации и является безальтернативной основой работы органов государственного и муниципального управления, обеспечивающих функционирование стратегических отраслей экономики [10].

Результаты исследования и их обсуждение

Разнообразие мира представляет собой огромное количество информации. Почти невозможно полностью отобразить все это,

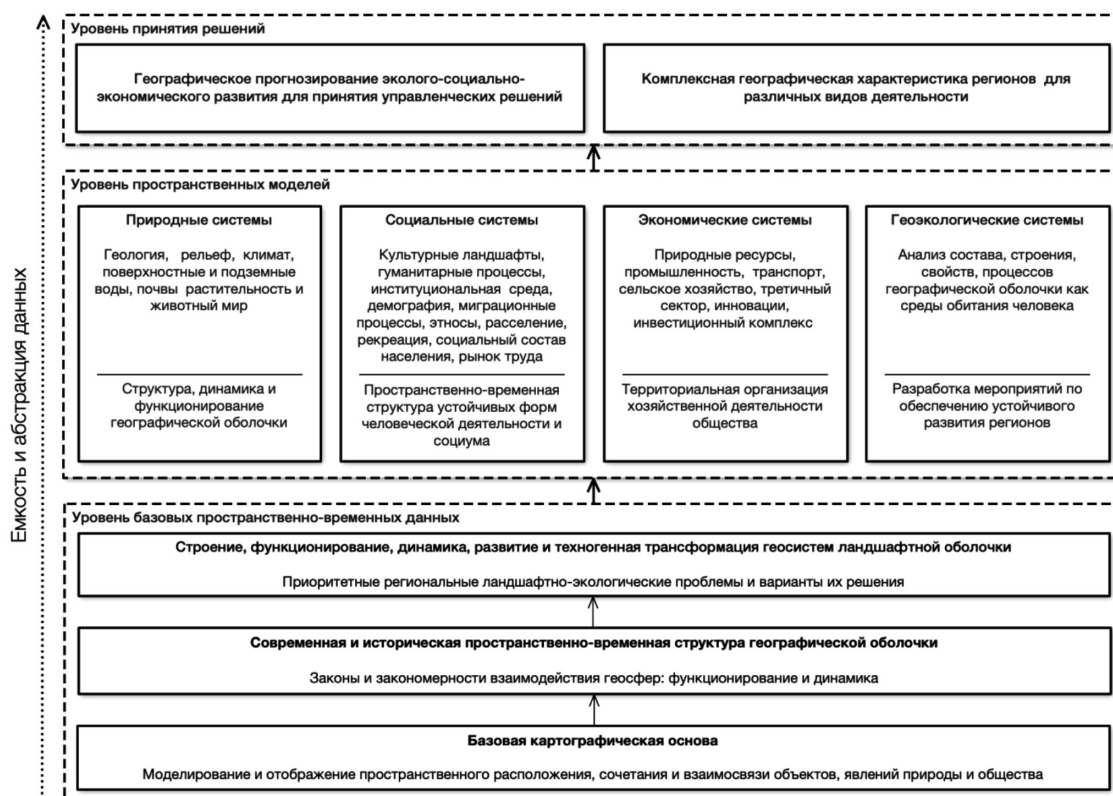
поэтому прибегают к определенному количеству обобщения и абстракции, то есть все богатство и разнообразие окружающего пространства приводит к созданию определенных моделей, то есть к конечному количеству некоторых природных объектов и свойств.

В результате экспериментов и первоначального обобщения материалов можно утверждать, что в ИПД для оценки процессов функционирования природно-социально-производственных систем целесообразно выделить следующие уровни интеграции и использования данных (рисунок):

а) *уровень базовых пространственно-временных данных*, включающий базовую картографическую основу и описывающий современную и историческую пространственно-временную структуру географической оболочки, закономерности взаимодействия геосфер;

б) *уровень пространственных моделей* предполагает систематизацию данных по следующим модулям:

1) «Природные системы» (описывает структуру, динамику и функционирование географической оболочки и её структурных частей; энергетические и информационные свойства геосфер);



Интегральный синтез геопространственной информации в цифровой ИПД

2) «Социальные системы» (систематизирует информацию о пространственно-временной структуре общества и устойчивых формах человеческой деятельности: культурные ландшафты, гуманитарные процессы, культура, институциональная среда);

3) «Экономические системы» (характеризует территориальную организацию хозяйственной деятельности общества: природные ресурсы, промышленность, транспорт, сельское хозяйство, третичный сектор, инновации, инвестиционный комплекс);

4) «Геоэкологические системы» (представляет пространственно-временные системы, отражающие структуру географической оболочки как среды обитания человека и ее изменение в условиях техногенеза и мероприятия по обеспечению устойчивого развития);

в) *уровень принятия решений* предполагает проектно-ориентированное использование данных для организации процессов географического прогнозирования (в первую очередь пространственно-временной прогноз эколого-социально-экономического развития) и формирования комплексной характеристики стран и регионов (систематизирующей данные о природе, населении, экономике, культуре и социальной организации).

Производство цифровых моделей в современных условиях представляет собой многоэтапный, относительно длительный процесс, при котором картографическая информация из графической формы преобразуется в цифровую форму, записывается на машиночитаемый носитель, подвергается повторной обработке и структурированию [14]. В этом процессе задействовано большое количество исполнителей разных специальностей, сложный набор аппаратного и программного обеспечения. В связи с этим возникновение ошибок при изготовлении цифровых моделей неизбежно, поскольку оно обусловлено сложностью системы и разнообразием факторов, которые на нее влияют. Очевидно, что основным показателем качества цифровых моделей является их достоверность – степень точности интеграции в системе картографической информации.

Для организации систем интеграции, хранения и обработки больших массивов пространственно-временной информации в цифровых ИПД целесообразно комплексное использование научно обоснованного набора систем управления базами данных (СУБД) [12]. Так, в реляционных хранилищах организация процесса интеграции данных основана на сущностях и связях, установленных между ними. Высокоорганизованная структура и гибкость делает

реляционные хранилища мощными и адаптируемыми к различным типам пространственных данных [13]. В цифровых ИПД целесообразно использование реляционных СУБД с поддержкой пространственных операций: PostgreSQL MySQL.

Альтернативные преимущества процессу интеграции пространственных данных способны предоставить NoSQL хранилища [14], которые следует классифицировать по следующим категориям: а) резидентные базы данных – Redis (предпочтительны для разработки систем кеширования данных, буферов высокоскоростного обмена с агентами Интернета вещей); б) документные базы данных – MongoDB, RethinkDB (актуальны для систематизации слабоструктурированных данных, информации мониторинга пространственных процессов, хранения и регистрация событий); в) графовые базы данных – Neo4j, JanusGraph (целесообразны для хранения информации о горизонтальных и вертикальных связях геосистем); г) колоночные базы данных – ClickHouse, Cassandra (представляют безальтернативное решение для организации модулей интерактивной аналитической обработки данных (OLAP-компонентов) в системе цифровых ИПД); д) базы данных временных рядов – InfluxDB, TimescaleDB (внедряются для сбора и управления пространственными данными, для которых характерно изменение с течением времени, в том числе показателей развития стихийных процессов и транзакций Интернета вещей).

Заключение

Сформулируем ключевые результаты исследования:

1. Современные цифровые методы анализа пространственной информации формируют основу исследования геосистем, представляющих иерархическую модель территории. Такая технология является перспективной, поскольку позволяет расширить возможности ИПД как систем для обработки и анализа пространственно-временной информации.

2. Эффективность анализа значительно возрастает при использовании хранилищ информации, в которой хранятся разнообразные данные, полученные из разных источников. Развитие функциональных возможностей ИПД является требованием современного научного подхода к оценке и анализу географической оболочки и ведет к более широким перспективам использования геоинформационных при поддержке принятия управленческих решений.

3. Ни одна отдельная парадигма к организации хранилища пространственно-вре-

менной информации в системе цифровых ИПД не может предоставить ключ к решению всех задач, вследствие необходимости различных представлений для разных типов пространственно-временных данных. Ответ на проблему интеграции, хранения и обработки больших массивов пространственно-временной информации в цифровых ИПД могут предоставить мультимодельные хранилища, гибридные базы данных, основанные на функциональных возможностях СУБД разных классов.

4. Следствием использования мультимодельных хранилищ являются усиление качественных характеристик цифровой ИПД – гибкое масштабирование хранилища по мере необходимости, повышение надежности и отказоустойчивости. Хранилище при этом может быть централизовано в центре обработки данных или децентрализовано в масштабах облака.

Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 20-37-70055.

Список литературы

1. Сочава В.Б. Введение в учение о геосистемах. Новосибирск: Наука, 1978. 320 с.
2. Николаев В.А. Классификация и мелкомасштабное картографирование ландшафтов. М.: МГУ, 1978. 62 с.
3. Космачев К.П. Проблемы экономико-географической экспертизы информационной базы территориальной организации производства // Доклады Института географии Сибири и Дальнего Востока. 1974. № 45. С. 3–11.
4. Солнцев Н.А. Природная география, ландшафтоведение и естествознание // Вестник Моск. ун-та, серия географ. 1977 № 1. С. 10–15.
5. Miklós L., Kočická E., Izakovičová Z., Kočický D., Špínerová A., Diviaková A., Miklášová V. Landscape as a geosystem. In Landscape as a geosystem. 2019. P. 11–42.
6. Yamashkin S.A., Yamashkin A.A., Zanozin V.V., Radovanovic M.M., Barmin A.N. Improving the Efficiency of Deep Learning Methods in Remote Sensing Data Analysis: Geosystem Approach. IEEE Access. 2020. № 8. P. 179516–179529.
7. Lu J., Holubová I. Multi-model databases: a new journey to handle the variety of data. ACM Computing Surveys (CSUR). 2019. No. 52 (3). P. 1–38.
8. Ямашкин А.А., Ямашкин С.А. Концепция устойчивого развития в региональном геопортале // Образование через всю жизнь для устойчивого развития: материалы Международной конференции. 2014. С. 535–541.
9. Ямашкин С.А. Структура регионального геопортала как инструмента публикации и распространения геопрограммированных данных // Научно-технический вестник Поволжья. 2015. № 6. С. 223–225.
10. Granell C., Miralles I., Rodríguez-Pupo L.E., González-Pérez A., Casteleyn S., Busetto L., Huerta J. Conceptual architecture and service-oriented implementation of a regional geoportal for rice monitoring. ISPRS International Journal of Geo-Information. 2017. No. 6 (7). 191 p.
11. Вдовин С.М., Федосин С.А., Ямашкин А.А., Ямашкин С.А. Получение, хранение и распространение геоданных как единый информационный процесс // Природные опасности: связь науки и практики: материалы II Международной научно-практической конференции. 2015. С. 82–90.
12. Taft R., Sharif I., Matei A., VanBenschoten N., Lewis J., Grieger T., Mattis P. Cockroachdb: The resilient geo-distributed SQL database. Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. 2020. P. 1493–1509.
13. Roumelis G., Vassilakopoulos M., Corral A., Manolopoulos Y. Efficient query processing on large spatial databases: a performance study. Journal of Systems and Software. 2017. Vol. 132. P. 165–185.
14. Han J., Haihong E., Le G., Du J. Survey on NoSQL database. 2011 6th international conference on pervasive computing and applications. 2011. P. 363–366.