DOI: 10.17117/nv.2015.04.053

Поступила (Received): 12.12.2015

<http://ucom.ru/doc/nv.2015.04.053.pdf>

kafedra_pi@rshu.ru

**Истомин Е.П., Колбина О.Н.,
Степанов С.Ю., Сидоренко А.Ю.**
**Анализ моделей и систем обработки гетерогенных
данных для использования в прикладных ГИС**

Istomin Y.P., Kolbin O.N., Stepanov S.Yu., Sidorenko A.Yu.
**Analysis of models and systems for processing
heterogeneous data for use in GIS applications**

В статье рассматриваются предложения по использованию моделей и функционирующих систем обработки гетерогенных данных. Представлено описание разработанных технологий и программных реализаций консолидации гетерогенной информации из различных источников

The article discusses suggestions for using models and operating systems for handling heterogeneous data. Presents a description of the developed technologies and software implementations consolidation of heterogeneous information from different sources

Ключевые слова: обработка данных, база данных, модель

Key words: data processing, database, model

Истомин Евгений Петрович

Доктор технических наук, профессор, декан
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский пр., 98

Istomin Eugene Petrovich

Doctor of Engineering Sciences, Professor, Dean
Russian State Hydrometeorological University
Saint-Petersburg, Malookhtinsky ave., 98

Колбина Ольга Николаевна

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский пр., 98

Kolbin Olga Nikolaevna

Senior lecturer
Russian State Hydrometeorological University
Saint-Petersburg, Malookhtinsky ave., 98

Степанов Сергей Юрьевич

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский пр., 98

Stepanov Sergey Yurievich

Senior lecturer
Russian State Hydrometeorological University
Saint-Petersburg, Malookhtinsky ave., 98

Сидоренко Артем Юсупович

Старший преподаватель
Российский государственный
гидрометеорологический университет
г. Санкт-Петербург, Малоохтинский пр., 98

Artem Sidorenko Yusupovich

Senior lecturer
Russian State Hydrometeorological University
Saint-Petersburg, Malookhtinsky ave., 98

Введение (Introduction)

Объединение гетерогенных данных для последующего совместного анализа и обработки становится популярной темой исследования последнего десятилетия. Основной задачей технологии слияния данных является объединение данных из разных источников в интересах решения последующих содержательных задач: принятие управленческих решений, анализ состояния объектов, прогнозирование ситуаций и т.д. Интеграция гетерогенных данных в единую информационную среду обеспечивает возможность их комплексного анализа и позволяет получить качественно новые знания об объекте исследования [1]. Такая информационная среда может объединять, следующие типы информационных ресурсов: геоинформационные ресурсы (картографические материалы, спутниковые снимки, данные наблюдений и т. п.); электронные каталоги; цифровые изображения, аудио и видео материалы; другие ресурсы.

В практическом плане увеличивается число комплексных программно-аппаратных решений. В задачи исследования ставится рассмотрение некоторых из функционирующих на данный момент систем обработки гетерогенных данных и других предложенных моделей. В центре внимания находится решение различных ситуативных прикладных задач [2].

В список моделей и систем обработки гетерогенных данных для использования в прикладных ГИС вошли следующие системы:

1. Распределенная геоаналитическая система «Geo-Status».
2. Сервис-ориентированная геоинформационная система для обеспечения доступа к пространственным данным в задачах исследования природных экосистем.

Материалы и методы (Materials and Methods)

Для достижения цели исследования вышеописанных систем, в работе были использованы специальные и общенаучные методы.

Стадии и этапы исследования:

1. Проведение анализа исследуемой геоаналитической системы.
2. Исследование процесса консолидации данных.
3. Анализ внутренней инфраструктуры геомодели.
4. Описание технологий доступа к гетерогенным данным.
5. Проведение анализа исследуемой сервис-ориентированной геоинформационной системы для обеспечения доступа к пространственным данным в задачах исследования природных экосистем.
6. Обзор структуры серверного программного обеспечения информационной системы.
7. Подведение результатов исследования.

Результаты (Results)

Распределенная геоаналитическая система «Geo-Status». Решение широкого спектра информационно-аналитических задач с привязкой к пространственным данным. Полнофункциональная информационная система с развитыми механизмами работы с базой данных. Работа с базой данных реализована

на платформе 1С. В качестве геосервера может использоваться как платный ArcGis Server, так и бесплатный GeoServer. Система многофункциональна, работает через Web-сервис, имеет развитые средства разделения доступа. В системе Geo-Status возможно оперативно, в реальном времени, получать необходимую информацию с визуальным отображением данных. Система позволяет в режиме реального времени, из любой точки обеспечить информационно-аналитическую поддержку процессов принятия управленческих решений.

На рис. 1 показана структурная схема системы.

Рис. 1. Структурная схема РГАС

Распределенная геоаналитическая система (РГАС) создана, как единый информационный ресурс. Ключевым элементом системы является ситуационный центр. РГАС обеспечивает информационную и аналитическую поддержку принятия эффективных управленческих решений на основе объективной и многопараметрической информации. РГАС создана на платформе пространственных данных, обеспечивает мониторинг приоритетных направлений развития и имеет широкую сферу применения. Структура распределенных данных в РГАС представлена на рис. 2.

Рис. 2. Слои представления данных в РГАС

Как видно из схемы, в состав общей базы данных ОБД Администрация города, входят другие базы данных, упорядоченные по отраслям и сферам применения. Для эффективного функционирования к системе РГАС предъявлены следующие требования:

- Автоматизированный процесс создания баз данных(сбор, хранение, анализ и передача информации);
- Хранение пространственных и атрибутивных данных для их дальнейшего анализа и обработки;
- Мониторинг в режиме реального времени (выполняется запрос о свойствах объектов и автоматизируется процесс анализа, сопоставляя множество параметров);
- Формирование и представление материалов для принятия управленческих решений.

Система реализована на 2 платформах.

– Первая платформа – 1С. Используется для организации работы с БД. За счет этого достигается максимальная концентрация функциональных возможностей при вводе, анализе и получении отчетности.

– Вторая платформа – геоинформационный сервис. Система поддерживает работу, как с платным, так и бесплатным сервером разработки «ESRI».

Система может работать, как через web-браузер, так и через специального тонкого клиента. Платформа 1С обеспечивает практически все современные методы межсистемного взаимодействия: com, soap, xml. Геоинформационный модуль базируется на открытой библиотеке «OpenLayers». Таким образом, возможна значительная адаптация функционала работы с картой. Геоинформационный модуль может работать не только с локальными слоями, но и использовать сторонние по открытым протоколам WMS, WFS. Интеграция РГАС с другими сервисами представлена на рис. 3.

Рис. 3. Интеграция с другими сервисами

Организация интеграционных процессов реализовано с помощью программного комплекса SONIC ESB. Комплекс предназначен для организации и

обеспечения информационного взаимодействия модулей ситуационного центра и взаимодействия с внешними по отношению к нему системами [3].

Особенности интеграции:

- Предоставление унифицированного интерфейса взаимодействия по общепринятым промышленным стандартам: SOAP, FTP, POP, JDBC.

- Интеллектуальная маршрутизация, настройка правил маршрутизации, основанных на атрибутах передачи данных, содержании данных и предопределенных настройках интеграционной платформы.

- Трансформация данных – предоставление данных в требуемом формате, обеспечение совместимости сопрягаемых систем и модулей за счет наличия специального интерфейса преобразования схем трансформации данных.

- Безопасное конфигурирование, управление и контроль функционирования всех сервисов трансформации данных.

Внедрение такой системы, предоставляет доступ к огромному объему данных, что способствует более эффективному принятию управленческих решений. Из описания системы можно сделать следующие интересующие по данной теме выводы:

- Система РГАС имеет доступ к различным базам данных – в свою очередь по отношению друг к другу эти БД гетерогенны.

- Для соединения с данными используются общепринятые промышленные стандарты: SOAP, FTP, POP, JDBC.

- Для использования в системе РГАС сторонних данных, необходимо учитывать правила предоставления информации в требуемом формате.

Сервис-ориентированная геоинформационная система для обеспечения доступа к пространственным данным в задачах исследования природных экосистем. В Институте вычислительных технологий СО РАН разрабатывается сервис-ориентированная геоинформационная система для обеспечения доступа к пространственным данным. Система строится на основе каталога спутниковых данных Новосибирского научного центра СО РАН, который включает архивные данные со спутников серии LandSat по территории РФ за 1982–2002 гг., и с 2008 г. регулярно пополняется оперативными данными SPOT 4. С апреля 2010 г. к каталогу подключена система структурного восстановления оперативных данных, поступающих с платформ TERRA/AQUA. Прием данных осуществляется Западно-Сибирским региональным центром приема и обработки данных в рамках соглашения о межведомственном сотрудничестве.

Система реализована в виде базового набора приложений, работающих в среде сервера приложений Tomcat. Подсистема пользовательских интерфейсов реализована с использованием технологий PHP/JavaScript. Доступ к системе реализован посредством модуля Central Authentication Service (CAS). Он позволяет организовать многоуровневую систему разграничения прав доступа с централизованной базой пользователей на основе LDAP-каталога Сибирского отделения РАН и реализовать практически индивидуальные настройки доступа к любому защищаемому ресурсу.

Система состоит из следующих функциональных блоков (рис. 4).

Рис. 4. Структура картографической информационной системы

Центральным блоком системы является подсистема картографических сервисов, реализованная на основе пакета GeoServer. Подсистема обеспечивает доступ к картографической информации, хранящейся в системе (базовые подложки, векторные слои, построенные по базам данных и др.). Для публикации динамических данных используется пакет UMN MapServer, который обеспечивает доступ к данным, формируемым в оперативном режиме, а также к пользовательским наборам данных.

Для тематической обработки данных в систему интегрирован комплекс программ, основанный на эффективных логико-вероятностных алгоритмах выбора информативных признаков и классификации. Для расширения функциональности системы используется подсистема сервисов. На основе WPS-сервера deegree, распространяемого по лицензии GPL, разработан модуль для интеграции в систему алгоритмов обработки пространственных данных. Он осуществляет интерпретацию входных и выходных данных согласно спецификации протокола WPS и выполняет функции контейнера для неограниченного числа WPS-процессов. Архитектура модуля представлена на рис. 5.

Для обеспечения функционирования системы в распределенном режиме и интероперабельности по протоколам доступа к метаданным и их представлению в нее интегрированы модули поддержки протокола Z39.50. Поисковая система позволяет не только находить данные по метаданным, но и выполнять комплексные запросы.

В настоящее время к системе подключено 26 институтов Сибирского отделения РАН. Система используется для выполнения междисциплинарных интеграционных проектов СО РАН. В систему внедрено несколько информационных ресурсов, разработанных сотрудниками Сибирского отделения РАН в рамках различных проектов и грантов и предоставляемых в виде веб-сервисов, в том числе:

Рис. 5. Структурная схема подсистемы сервисов

- векторная карта растительности Западной Сибири и ландшафтная карта Иркутской области;
- векторная карта почв бореальной зоны Западно-Сибирской равнины и соответствующая ей карта растительности, содержащая 28 различных типов растительности (М 1 : 7 500 000);
- база данных по содержанию органического углерода в почвах Сибири;
- климатологические данные за период с 1989 по 2009 г. на ключевой участок с координатами: 53–70° с. ш., 59–93° в. д., подготовленные сотрудниками Института мониторинга климата и экологических систем СО РАН.

Разработан подход к интеграции разнородных пространственных данных, позволяющий осуществлять их комплексный анализ. Для этого все необходимые данные приводятся к единому пространственному разрешению с использованием регулярной сетки. Разработанный подход применялся для построения информационных моделей бореальных экосистем Западной и Восточной Сибири.

Если для интеграции спутниковых и картографических данных для минимальной функциональности информационной системы достаточно интеграции в рамках, например, каталога GeoNetwork, то для более широкого спектра информационных ресурсов функциональных возможностей этой системы становится явно недостаточно. Несмотря на то, что GeoNetwork при загрузке дополнительных схем данных и надлежащей настройки поддерживает каталогизацию и обеспечение доступа к ресурсам различного типа, способы управления ресурсами в этой системе оставляют желать лучшего. В этой системе отсутствует поддержка иерархических коллекций, включающих в том числе и разнородные информационные ресурсы, а также детализированное разграничение доступа к этим коллекциям на основе расширенных ролевых правил. Отсут-

ствие этой функциональности в GeoNetwork существенно сужает рамки ее использования для коллективной работы по созданию тематической информационной системы.

В качестве основы системы, интегрирующей разнородные данные, была выбрана система DSpace. Информационная система DSpace обладает широкими возможностями по управлению цифровым контентом, но не содержит интерфейсов для работы с географическими координатами. Учитывая, что DSpace широко используется для создания электронных библиотек, мы не могли не модифицировать эту систему для придания ей дополнительной функциональности.

Рис. 6. Структура серверного программного обеспечения информационной системы

На рис. 6 показана общая схема информационной системы, реализующей электронную библиотеку. Наряду с компонентой, управляющей цифровым контентом (DSpace), система включает СУБД PostgreSQL для хранения метаданных, LDAP-сервер для обеспечения функции однократной аутентификации в корпоративном каталоге СО РАН, а также сервер ZooPARK для обеспечения географического поиска в ресурсах DSpace и обеспечения интеграции с другими информационными ресурсами, в том числе и с ресурсами ГИС [4].

Обсуждение и заключение (Discussion and Conclusion)

На основе анализа представленных ГИС систем и моделей, можно сделать очевидный вывод, что раздельно существуют системы аналитической обработки и оперативного управления данными, системы управления документами и прочие системы. Различные задачи требуют использования различных моделей представления данных. На этом фоне можно поставить задачу интеграции гетерогенных данных, которую можно было бы рассматривать как основу для реализации системы, поддерживающей оперативное управление разнородными данными и их аналитическую обработку. Ряд аспектов интеграции данных ГИС рассмотрен в публикациях [8-11].

Разработка технологии слияния гетерогенной информации из различных источников требует получения ответов на следующие вопросы:

- каким образом может быть решена проблема консолидации данных, полученных из распределенных гетерогенных источников;
- какой должна быть архитектура системы слияния данных.

Практика решения задач устойчивого развития территориями показывает, что они имеют комплексный характер и требуют привлечения большого количества разнообразной информации, получаемой из разных источников. В наибольшей степени при этом используются геопространственные данные, необходимы также и другие данные, без которых решение задач невозможно или недостоверно. Комплексное использование такой информации связано с большими проблемами, обусловленными, прежде всего, гетерогенностью данных различной природы. Использование геопространственных данных связано с необходимостью постоянно интегрировать эти данные с геопространственными данными другой природы. При этом большая часть проблем связана с реализацией механизмов интегрирования и преодоления гетерогенности данных.

В постановке задачи слияния данных можно выделить две основные задачи, имеющие большое значение:

- Развитие модели доступа к распределенным гетерогенным данным;
- Определение единого формата представления данных.

Предложенные в данной работе общие структуры моделей интегрирования гетерогенных данных могут быть взяты за основу при создании подобных моделей в других областях применения пространственных данных.

Список используемых источников:

1. Колбина О.Н., Зоринова Е.М., Истомин Е.П., Степанов С.Ю. Обзор технологических методов доступа к базам данных распределенных ГИС // Информационные технологии и системы: управление, экономика, транспорт, право. Сб. тр. Международной практической конференции «Инфогео-2014». Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
2. Степанов С.Ю., Гришин Н.М., Колбина О.Н., Соловьева А.Д. ГИС. Обзор программного обеспечения // Информационные технологии и системы: управление, экономика, транспорт, право: Сб. тр. Международной практической конференции «Инфогео-2014». Вып. 3(14). СПб: Андреевский издательский дом, 2014. 116 с.
3. Управление федеральной службы государственной регистрации, кадастра и картографии – информационный портал.
URL: http://www.to33.rosreestr.ru/upload/to33/files/Конценция_ситуационного_центра_АВТ
4. Жижимов О.Л., Молородов Ю.И., Пестунов И.А., Смирнов В.В., Федотов А.М. Интеграция разнородных данных в задачах исследования природных экосистем. Вестник НГУ. Серия: Информационные технологии. 2011. Том 9. Вып. 1.
5. Шокин Ю.И. Геоинформационные технологии и математические модели для мониторинга ГЗС и управления экологическими и социально-экономическими системами; под ред. И.Н. Ротановой; Рос. акад. наук, Сиб. отделение, Ин-т водных и экологич. проблем. Барнаул: Пять плюс, 2011. 250 с.
6. Ильин Е.В. Технологическая платформа для распределенных гетерогенных систем сбора и обработки данных // Материалы 50-й Международной научной студенческой конференции «Студент и научно-технический прогресс»: Информационные технологии. Новосибирск, 2012.
7. Степанов С.Ю. Сравнительный анализ открытых геоинформационных систем // Информационные технологии системы: управление, экономика, транспорт, право. СПб: Андреевский издательский дом, 2013. 156 с.
8. Котиков П.Е. Варианты построения темпоральных баз данных в геоинформационных системах // Научный аспект. 2014. № 4. С. 118-120.

9. Котиков П.Е., Муратова О.Н. Существующие сложности перехода к применению пространств данных в геоинформационных системах // Актуальные проблемы технических наук. Уфа: Аэтерна, 2015. С. 42-44.
10. Котиков П.Е. Некоторые решения технологических проблем хранения пространственных данных ГИС // Научный аспект. 2015. № 2. С. 212-214.
11. Котиков П.Е., Муратова О.Н. Обзор вариантов архитектурных решений построения распределенных и параллельных баз данных для геоинформационных систем // Закономерности и тенденции развития науки : Сборник статей Международной научно-практической конференции. Издательство «Аэтерна». Уфа, 2015. С. 52-54.

References:

1. Kolbin O.N., Zharinova E.M., Istomin Y.P., Stepanov S.Yu Review of technological methods of access to databases distributed GIS // Information technology and systems: management, economy, transport, right: International practical conference "Infogeo-2014". Vol. 3(14). SPb: Publishing house Andriivskiy, 2014. 116 p.
2. Stepanov S.Y., Grishin N.M. Kolbin O.N., Soloveva A.D. GIS. Software review // Information technology and systems: management, economy, transport, right. International practical conference "Infogeo-2014". Vol. 3(14). SPb: Publishing house Andriivskiy, 2014. 116 p.
3. The Federal service of state registration, cadastre and cartography and information portal. URL: http://www.to33.rosreestr.ru/upload/to33/files/Концепция_ситуационного_центра_ABT
4. Zhizhimov O.L., Molorodov Yu.I., Pestunov I.A., Smirnov V.V., Fedotov A.M. Integration of heterogeneous data in study of natural ecosystems. Vestnik NSU. Series: Information technologies. 2011. Vol. 9. Issue 1.
5. Shokin Y.I. Geoinformation technologies and mathematical models for monitoring G and management of environmental and socio-economic systems; under the editorship of I.N. Romanovoj; ROS. Acad. Sciences, Sib. Department, Institute of water and environmental. problems. Barnaul: Five plus 2011. 250 p.
6. Ilyin E.V. Technology platform for distributed heterogeneous systems of data collection and processing. Proceedings of the 50th International scientific student conference "Student and scientific-technical progress": Information technologies. Novosibirsk, 2012.
7. Stepanov S.Y. Comparative analysis of open geographic information systems/Information technology systems: management, economy, transport, right. St. Petersburg: Andrew Izdatelsky house, 2013.156 p.
8. Kotikov P.E. Options for the design of temporal databases in geoinformacionnyh systems. Scientific aspect. No. 4. 2014. P. 118-120.
9. Kotikov P.E., Muratova O.N. The Existing difficulties of the transition to the use of data spaces in geographic information systems // Actual problems of technical Sciences : materials of International scientific-practical conference. Ufa, 2015. P. 42-44.
10. Kotikov P.E. Some of the technological problems of spatial data storage GIS / Scientific aspect No. 2. 2015. P. 212-214.
11. Kotikov P.E., Muratova O.N. Overview of architectural solutions for building distributed and parallel data bases for geographic information systems // Regularities and tendencies of development of science: collected papers of International scientific-practical conference. Ufa, 2015. P. 52-54.

© 2015, Истомин Е.П., Колбина О.Н., Степанов С.Ю., Сидоренко А.Ю.

Анализ моделей и систем обработки гетерогенных данных для использования в прикладных ГИС

© 2015, Istomin Y.P., Kolbin O.N., Stepanov S.Yu., Sidorenko A.Yu.

Analysis of models and systems for processing heterogeneous data for use in GIS applications