

Сервисы геоинформационной системы сбора, хранения и обработки данных натурных наблюдений ♣

© Молородов Ю.И., Смирнов В.В., Федотов А.М.

Институт вычислительных технологий СО РАН
yumo@ict.nsc.ru

Аннотация

Описаны сервисы, расширяющие возможности геоинформационного сервера построенного на основе стандартов Open Geospatial Consortium. Предлагается модель распределенной информационно-аналитической системы с единой точкой доступа к геоданным (через разнородные пространственно распределенные базы метаданных) и к инструментарию для их обработки и визуализации.

1 Введение

Создание информационных ресурсов и интеграция их в единую информационную среду являются приоритетными направлениями развития современного общества. Разработка механизмов, обеспечивающих как функционирование общей информационно-аналитической среды, так и доступ к научным ресурсам, и их сохранность, имеет первостепенное значение в задачах информационной поддержки научных исследований. Эти вопросы особенно важны при исследованиях в области экологии и, в частности, биологического разнообразия, выполняемых в рамках междисциплинарных проектов СО РАН. Исследования проводятся различными группами ученых, разделенными географически, связанных необходимостью проведения совместных работ, обмену данными и координации своих действий. Применение информационных технологий в науках о Земле способствует пониманию как глобальных, так и региональных природных процессов, формирующих окружающую среду.

Развивающиеся геоинформационные технологии в настоящее время переходят на новый этап развития – создание распределенных ГИС (РГИС). Предпосылками этого перехода стали внедрение передовых технологий (распределенные базы данных, распределенные вычисления, высокоскоростные линии связи) и доступность веб-технологий широкому кругу пользователей. Этому

также способствует стремительный прогресс в области создания и развития средств и технологий дистанционного зондирования Земли, обеспечивающий миллионам пользователей доступ к данным, получаемым с помощью спутниковых систем нового поколения (QuickBird, IKONOS, OrbView, GeoEye, SPOT, TerraSarX, Ресурс-ДК и др.) [1]. С другой стороны, в последние годы крупнейшие разработчики программного обеспечения объединяются в консорциумы по стандартизации. В области геоинформатики создан Консорциум открытых ГИС (Open Geospatial Consortium – OGC) [2]. Соответствие стандартам OGC позволяет программным продуктам различных производителей (в том числе коммерческим) взаимодействовать при решении конкретных задач. Переходу к РГИС также способствует внедрение спутниковых систем навигации NAVSTAR GPS и ГЛОНАСС, позволяющих с высокой точностью определять местоположение объекта в любой точке планеты. Распределенные геоинформационные системы имеют неоспоримые преимущества перед настольными, благодаря таким факторам, как:

- распределенный доступ к системе (наличие веб-интерфейса, позволяющего избежать установки дорогостоящего программного обеспечения, простота изменения и обновления программного обеспечения, доступность широкому кругу пользователей);
- распределенное хранение данных (организация доступа к архивам данных, возможность хранения пользовательских данных на сервере);
- распределенная обработка данных (возможность проведения обработки на высокопроизводительных вычислительных системах).

В настоящей работе представлена модель информационно-аналитического портала с единой точкой доступа к широкому диапазону данных и к инструментарию для их визуализации и обработки на высокопроизводительных вычислительных системах. Потребность в системах подобного рода особенно остро ощущается при проведении фундаментальных и прикладных исследований в области экологии и рационального природопользования.

2 Принципы построения и структура Системы

Разрабатываемая геоинформационная система базируется на операционных системах семейства UNIX и наборе специализированных программных продуктов с открытым исходным кодом, распространяемых под лицензией GPL (GNU General Public License). Она полностью удовлетворяет требованиям OGC, предъявляемым к аналогичным системам, и допускает возможность подключения других ГИС. Основная цель ее разработки – создание виртуальной информационно-аналитической среды для поиска, обработки и анализа пространственных данных. Она позволит организовать единую точку доступа к различным геоинформационным системам и распределенным хранилищам данных и атрибутивной информации.



Рис. 1. Структура распределенной ГИС

На рис. 1 представлена структура системы. Доступ осуществляется через стандартный веб-браузер, что обеспечивает платформенную независимость. Многоуровневая система разграничения прав доступа после авторизации позволяет генерировать “на лету” графический интерфейс пользователя (в соответствии с уровнем доступа). Он представляет собой виртуальный рабочий стол с традиционными элементами управления. Ядро виртуальной среды состоит из набора Perl/PHP/JavaScript-приложений (с поддержкой технологии AJAX) и интерфейсов к внешним компонентам системы, работающих под управлением веб-сервера Apache. Внешние приложения взаимодействуют с системой через интерфейсы, описанные на языке XML.

Apache 2.0 (с расширением Tomcat) используется в качестве *HTTP-сервера* для платформы UNIX. Это веб-сервер разрабатывается и поддерживается открытым сообществом разработчиков под эгидой Apache Software Foundation и включен во многие программные продукты, среди которых СУБД Oracle и IBM WebSphere. К его основным достоинствам относятся надежность и гибкость конфигурации. Для размещения пространственно-координированных данных используется *картографический сервер* UMN MapServer. В

настоящее время он приобретает все большую популярность среди разработчиков геоинформационных веб-сервисов. По своим функциональным возможностям он не уступает коммерческим программам, а по ряду параметров превосходит их, в частности по производительности [10]. MapServer включает в себя все необходимое для разработки основных картографических сервисов WMS/WFS, рекомендованных к использованию OGC [11]. Он позволяет пользователю формировать карты с одновременным использованием материалов, хранящихся как в локальных, так и в удаленных архивах. Богатая функциональность, легкость интеграции с различными СУБД и открытость исходных кодов предопределили популярность программы.

В качестве базового *инструментария* для *обработки и анализа данных дистанционного зондирования* выбран пакет программ GRASS GIS (Geographic Resources Analysis Support System). Его отличительной особенностью является интеграция в среду UNIX, поддержка основных типов пространственных данных, мощный процессор обработки растровых данных, модульность и открытый инструментарий для быстрой и эффективной разработки модулей расширения. Использование в пакете библиотек GDAL и PROJ4 обеспечивает поддержку всех современных стандартов геоданных и большой набор функций для трансформации и перепроецирования изображений. Пакет GRASS GIS позволяет разрабатывать модули расширения практически на всех языках программирования, для которых есть компилятор под UNIX (Perl, sh, C/C++, Fortran и т.д.). Он включает библиотеки для работы практически со всеми современными СУБД. Внутреннее представление растровых данных базируется на вокселях (3D-пикселях), что существенно повышает возможности по обработке. Интеграция пакета GRASS GIS в разрабатываемую систему позволяет с минимальными временными затратами обеспечить пользователя доступом к полнофункциональной ГИС, расширенной специализированным математическим аппаратом.

Большое внимание в системе уделено статистической обработке данных и визуализации результатов. Для этих целей в нее интегрирован статистический пакет R., распространяемый на условиях лицензии GPL.

Функциональность системы расширяется за счет использования *сервера приложений*. С обеспечения работы дополнительных сервисов, на языке XML, разработаны интерфейсы для взаимодействия с внешними приложениями.

Функционирование предлагаемой системы в распределенном режиме по протоколам доступа к метаданным основано на *модулях поддержки протокола Z39.50* [3] (и CIP [4] как одного из его профилей). Для этого в систему включены:

- сервер Z39.50 (ZooPARK), обеспечивающий базовую функциональность сервисов Z39.50 в

соответствии с различными прикладными профилями;

- шлюз Z39.50-HTTP, обеспечивающий простые пользовательские интерфейсы для доступа к ресурсам Z39.50 по протоколу HTTP;
- набор динамических провайдеров данных, каждый из которых описывает условия и протокол взаимодействия с конкретной СУБД, в которой хранятся метаданные. Вся логика работы с конкретной СУБД локализована в соответствующем провайдере данных. Взаимодействие базового сервера с провайдерами данных осуществляется через единый интерфейс.

2.1 Каталог пространственных данных

Для добавления данных дистанционного зондирования в хранилище системы создан специализированный сервис. Он производит начальную обработку данных, с последующей индексацией. После обработки данные перемещаются для долгосрочного хранения в систему хранения данных, а полученные метаданные приводятся к стандартному виду и размещаются в соответствующих разделах поисковой системы. Только после этого данные становятся общедоступными. Разработанная поисковая система позволяет находить данные по метаданным и выполнять комплексные запросы (содержащие географические координаты, номера трека и кадра, дату и время, параметры облачности и др.).

Для систематизации данных, организации поиска и извлечения из архива необходимой информации разработан каталог пространственных данных. В его основе лежит набор стандартных и специализированных программных продуктов с открытым исходным кодом, распространяемых под лицензией GPL (GNU General Public License). В основе структуры каталога лежит набор Perl/Java/JavaScript-приложений, работающих под управлением веб-сервера Apache.

Доступ к каталогу реализован посредством модуля Central Authentication Service (CAS), разрабатываемого в рамках проектом JA-SIG [5]. Он позволяет организовать многоуровневую систему разграничения прав доступа с централизованной базой пользователей на основе LDAP-каталога Сибирского отделения РАН. Функционально сервис реализован в виде прокси-сервера для аутентификации, возможна тонкая настройка параметров доступа (например, в зависимости от IP-адреса, доменного имени, даты и времени суток, используемого метода аутентификации, расширения файла и т.п.). Это позволяет реализовать практически индивидуальные настройки доступа к любому защищаемому ресурсу. Модуль CAS позволяет легко создавать защищенные ресурсы как на основе Apache/Tomcat, так и при использовании технологий PHP/JavaScript на платформе Apache.

2.2 Картографические сервисы

GeoServer предназначен для публикации набора векторных и растровых слоев, в частности слоя покрытия территории данными оперативного спутникового мониторинга за интересующий период. Приложение взаимодействует непосредственно с СУБД PostgreSQL/PostGIS что позволяет построить высокопроизводительный и легкий в настройке сервис.

Подсистема картографических сервисов реализована на программных продуктах, распространяемых под лицензией GPL (GeoServer и UMNMapServer).

Картографический сервер UMN MapServer предназначен для публикации растровых данных, и для извлечения информации из сопряженных баз данных и публикации законченных проектов. В настоящее время он становится одним из наиболее популярных инструментов для создания геоинформационных веб-сервисов и ресурсов[7]. MapServer содержит все необходимое для разработки картографических сервисов WMS/WFS, в соответствии со спецификациями OGC. Он позволяет формировать карты, одновременно используя информационные слои, размещенные как в локальных, так и в удаленных архивах. Богатая функциональность, легкое взаимодействие с различными СУБД и открытость исходного кода предопределили популярность программы.

К настоящему времени запущен в эксплуатацию картографический сервис, позволяющий организовывать хранение и публикацию в сети Интернет различных картографических материалов. Он обеспечивает доступ к сопряженной атрибутивной информации. Сервис обеспечивает доступ к картографическим материалам посредством стандартизированных протоколов (WMS, WFS), что позволяет использовать данный сервис как для разработки различных веб-приложений, так и для использования данных при работе в различных геоинформационных продуктах (например, ENVI, ArcGIS, MapINFO и др.).

2.2.1 Сервис доступа к картографическим данным

Картографический сервис обеспечивает доступ к картографическим данным посредством использования стандартов WMS/WFS и решает задачи:

1. обеспечения централизованного доступа к картографическим сервисам для организаций и сотрудников СО РАН;
2. повышение скорости работы с удаленными картографическими сервисами;
3. позволяет кэшировать поток данных предоставляемых сервисами, для снижения нагрузки на сервера, выполняющие функции генерации данных;
4. снижает загрузку внешних каналов связи.

В основе сервиса - система хранения данных ИБТ СО РАН с общим объемом дисковой памяти около 40 тбайт (рис. 2).

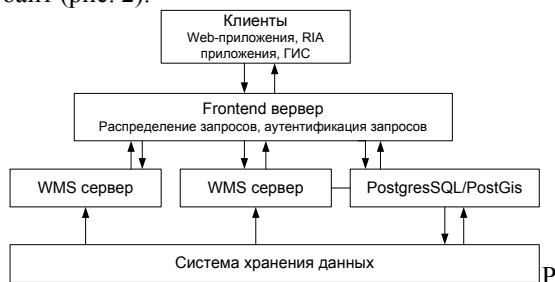


рис.2. Структурная схема сервиса

Программная реализация серверной части системы выполнена на основе картографического сервера *GEOSERVER*, распространяемого на основе лицензии с открытым исходным кодом (GPL). Сервер обеспечивает доступ пользователей посредством использования протоколов WMS/WFS, с использованием, как программный клиентов реализованных с использованием веб-технологий, так и с использованием стандартного, в том числе коммерческого (ArcGIS, ENVI и др.), программного обеспечения. Хранение атрибутивной информации, а также связанной пространственной информации реализовано на платформе PostgreSQL/PostGIS, распространяемой по лицензии GPL. Использование данной СУБД позволяет легко развернуть полномасштабную систему с возможностью работы с пространственной информацией.

Важной составляющей в разрабатываемой системе является подсистема кэширования геопространственных данных. Использование технологий кэширования позволяет существенно увеличить время доступа к данным, а также значительно снизить нагрузку на серверное оборудование. Дополнительно, при использовании технологий обмена содержимым кэша, имеется возможность строить распределенные по регионам системы кэширования, обеспечивающие оптимальный доступ пользователей к данным. Подсистема кэширования пространственных данных предполагает развитие нескольких стратегий работы с пользователями сервиса:

1. Организация кэширования во время выполнения текущих запросов.
2. Кэширование дополнительных областей, вокруг зон повышенного интереса.
3. Постепенное кэширование зон интересов по заказу.

Пространственный поиск по картографическим данным обеспечены протоколами Z39.50 и SOAP.

2.2.2 Векторная карта ландшафтов юга Восточной Сибири

На платформе Flash/Flex реализовано клиентское приложение WMS Browser, которое обеспечивает работу с различными картографическими слоями. Через это приложение реализован интерфейс работы с векторной картой ландшафтов

растительности юга Восточной Сибири (Иркутская область и Республика Бурятия, масштаб 1:1 500 000). Авторы карты: В.Б. Сочава, В.С. Михеев, О.П. Космакова (цифровая версия карты представлена И.Н. Владимировым и А.А. Сороковым). Адрес ресурса (<http://gis-app.ict.nsc.ru/wmsbrowser>). Здесь разными цветами выделены группы и геомы растительности.

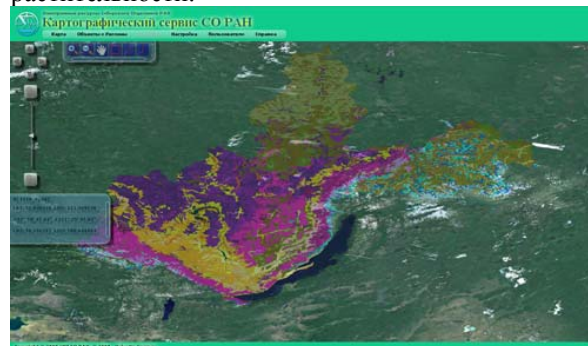


Рис.3. Приложение WMSBrowser

2.2.3 Электронные библиотеки для ботаники

Одной из важных задач создания геоинформационного портала (<http://gis-app.ict.nsc.ru>), является создание виртуальной информационно-аналитической среды для поиска, обработки и анализа спутниковых и наземных данных [8].

В Новосибирском научном центре специалисты ботаники Центрального сибирского ботанического сада (ЦСБС) и специалисты по информационным технологиям Института вычислительных технологий (ИВТ) СО РАН, начиная с 1993 г. занимаются разработкой программных продуктов и информационных ресурсов для нужд биологических наук. Основными направлениями разработок были выбраны: 1) создание тематических виртуальных библиотек - публикация обобщенной и систематизированной биологической информации с возможностью оперативного поиска; 2) разработка виртуальных коллекций - программных продуктов для хранения и систематизации первичных биологических данных. Были созданы Электронный атлас "Биоразнообразие животного и растительного мира Сибири", построена база данных "Зеленая книга Сибири", электронная библиотека "Разнообразие растительного и животного мира Сибири" и "Электронный каталог растений Сибири". Практически они являлись электронными версиями опубликованных в издательстве Сибирского отделения "Наука" книг Зеленой книги Сибири и Флора Сибири.

Они представляют собой виртуальные путеводители (электронные библиотеки) по флоре Сибири и Дальнего Востока России. Здесь можно узнать о распространении отдельных родов и видов в разных регионах Сибири и Дальнего Востока, в каких местообитаниях они чаще всего встречаются, посмотреть фотографии и рисунки, познакомиться с

общим распространением видов. Материалы, представленные в этих разработках, преследовали, по крайней мере, две цели: изучение возможностей современных вычислительных технологий для организации многопользовательской информационной системы по растениям крупного региона Сибири и определение эффективности работы с электронным информационным ресурсом по сравнению с публикацией на бумажных носителях. Главным отличием электронных библиотек стала возможность размещения в них многочисленных цветных иллюстраций и фотографий растений, выполненных в природной обстановке.

Кроме публикации обобщенной информации, насущной проблемой биологических (ботанических в частности) исследований является формализация, интеграция и систематизация первичных описательных данных; в противном случае накопление информации быстро утрачивает смысл. На ГИС-портале ИВТ СО РАН реализована возможность создания многопользовательских виртуальных коллекций. Предоставляемые в виртуальных коллекциях возможности отвечают основным потребностям первичной обработки ботанической информации. Помимо чисто научного значения, построение и развитие виртуальных коллекций важно для природоохранной деятельности – это позволяет оперативно выявлять действительно редкие виды, анализировать их распространение, выявлять наиболее интересные и подлежащие охране территории.

На страницах портала реализован доступ к информационным ресурсам и разработанным ранее базам данных: (<http://old.ict.nsc.ru/win/elbib/bio/>) "Биоразнообразие животного и растительного мира Сибири", (<http://old.ict.nsc.ru/win/elbib/bio/green/>) - "Зеленая книга Сибири" электронной библиотеки "Разнообразие растительного и животного мира Сибири" и Электронный каталог растений Сибири (<http://old.ict.nsc.ru/win/elbib/atlas/flora/>).

2.2.4 Атлас «Мхи России»

Создан прототип информационно-поисковой системы для хранения связанной и формализованной информации о видах растений (<http://gis-app.ict.nsc.ru/bio/>). В качестве исходного массива данных используется бриологическая база данных "Мхи России" (www.arctoa.ru), которая была получена путем объединения личных БД специалистов-бриологов России. Информационная система содержит базу данных "Мхи России", которая на текущий момент содержит около 30 000 документов.

Атлас построен по модульному принципу. Существует ядро, включающее в себя несколько программ, и модули, которые вызываются ядром по мере необходимости. Подобный подход позволяет уменьшить расходы, как связанные с модификацией системы в целом, так и отдельных её частей. Следует отметить, что в данном случае *модульный*

подход используется как принцип построения системы в целом, а не отдельных её частей. Внутри каждого модуля и ядра может использоваться как модульный принцип программирования, так и объектно-ориентированный.

Среди главных подсистем ядра можно выделить следующие блоки:

- разграничения доступа и управления пользователями;
- изменения фактографической информации;
- поиска информации и вывода результатов;
- взаимодействия с географической картой;
- многоязычного интерфейса.

При этом каждый из них представляет собой один или несколько программных модулей.

База данных состоит из трех основных элементов:

1. Таблица названий видов – "Species_list".

Содержит названия видов мхов, авторов названий, согласно сводке (Ignatov, Afonina, Ignatova et al., 2006), синонимы и русские названия. Возможно редактирование названий соответственно современной номенклатуре..

2. Таблица регионов России – "Regions_Russia".

Содержит данные о регионах России: код региона, буквенный код региона, русское и английское название (разработаны М.С.Игнатовым (ГБС РАН, Москва) для «Флоры мхов России» (www.arctoa.ru)).

3. Таблица коллекции Гербарных Этикеток – "etic_Russia".

Содержит в себе данные о конкретных образцах, взятых при полевых исследованиях.

Важным этапом в работе являлось обеспечение двустороннего взаимодействия с картой. Работа подсистемы осуществляется следующим образом. После выдачи поискового запроса, пользователь имеет возможность перейти по ссылке «показать все результаты поиска на карте». При активизации ссылки (нажатии на нее) программа находит в базе данных координаты всех найденных образцов и записывает их во временный файл специального вида. После этого открывается картографический интерфейс созданный файл передается программескрипту отображения. Далее он добавляет данные из этого временного файла отдельным слоем на карту, происходит центрирование и изменение масштаба для комфортного просмотра результатов (Рис.4).

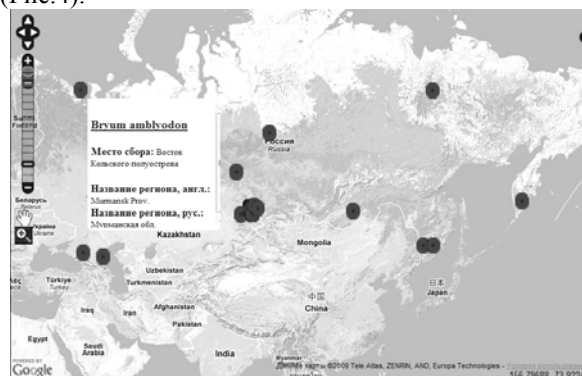


Рис. 3. Страница отображения гербарных этикеток на карте

Взаимодействие с картой происходит в обе стороны, так после просмотра результатов на карте, пользователь может перейти к просмотру конкретной гербарной этикетки. После щелчка по соответствующему маркеру всплывает краткое описание этикетки, на котором присутствует ссылка на полное описание. Либо как уже отмечалось выше выбрать с помощью картографического интерфейса область поиска данных

3. Заключение

В работе описаны сервисы распределенной информационно-аналитической системы, обеспечивающей поддержку исследований, связанных с обработкой и анализом пространственно распределенных данных. Создаваемая система взаимодействует с системой хранения данных общим объемом памяти более 40 Тбайт. Функциональные блоки объединяются в единую систему и производится разработка графического интерфейса пользователя и подключение баз данных, созданных в ходе многолетних исследований специалистами СО РАН. Создан прототип информационно-поисковой системы для хранения связанной и формализованной информации о видах растений (<http://gis-app.ict.nsc.ru/bio/>). В качестве исходного массива данных используется брiологическая база данных "Мхи России", которая была получена путем объединения личных БД специалистов-брiологов России. Первоначальный вариант базы данных был подготовлен в формате MS Access, что потребовало разработки алгоритмов преобразования в формат базы данных MySQL. Информационная система содержит базу данных "Мхи России", которая на текущий момент содержит около 30 000 документов.

Поиск в базе данных нужного вида мха осуществляется как по метаданным гербарному или полевому номеру образца, по виду и внутривидовому таксону, по точке сбора и местообитанию и др., так и по географическим параметрам и высоте. Результаты поиска могут быть выведены в таблице или в виде маркера на географической карте. При наведении на него курсора мыши, высвечивается гербарная этикетка с указанием названия вида, региона и места сбора.

Информационная система используется отечественными и зарубежными специалистами-брiологами. Она является основой для обобщения и ревизии гербарных материалов к подготовке первого издания «Флоры мхов России».

Литература

[1] SOVZOND Материалы Междунар. конф. "Космическая съемка - на пике высоких технологий". Россия, Москва, 18-20 апреля 2007 г. // <http://www.sovzondconference.ru/archive-2007/rus/agenda.html>.

[2] The Open Geospatial Consortium, 2007. <http://www.opengeospatial.org>.
 [3] ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50. 1995.
 [4] Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. Новосибирск: ОИГТМ СО РАН; ИВТ СО РАН, 2004. 361 с.
 [5] Catalogue Interoperability Protocol (CIP) Specification - Release B 2.4. CEOS/WGISS/ PTT http://www.dfd.dlr.de/ftp/pub/CIP/_documents/cip2.4/S_cover.pdf.
 [6] <http://www.ja-sig.org/products/cas/index.html>.
 [7] ANSI/NISO Z39.50-1995. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. Z39.50 Maintenance Agency Official Text for Z39.50. 1995
 [8] Шокин Ю.И., Добрецов Н.Н., Пестунов И.А., Молородов Ю.И., Смирнов В.В., Сиявский Ю.Н. Система сбора, хранения и обработки спутниковых и наземных данных Новосибирского научного центра СО РАН // Выч. тех. – 2008. – Т.13. – Вестн. КазНУ им. аль-Фараби. Серия: Математика, механика, информатика. – №4 (59). – Совместный вып. по материалам междунар. конф. «Вычислительные и информационные технологии в науке, технике и образовании». – Ч.III. – С.371-376.

Services GIS data collection, storage and data processing field observation

Molodtsov Yu.I., Smirnov V.V., Fedotov A.M.

We describe the services that enhance the possibility of geo-information server based on the standards of Open Geospatial Consortium. The model of distributed data-processing system with a single point of access to geo-data (in a spatially heterogeneous distributed database of metadata) and tools for processing and visualization.

* Работа частично финансировалась Президентской программой НШ 931.2008.9, РФФИ (грант № 09-07-00277), и Интеграционным проектом СО РАН № 50.