

Портал интеграции пространственных данных «ГеоМета»*

© О.М. Атаева, К.А. Кузнецов, В.А. Серебряков, В.И. Филиппов

Вычислительный центр им. А.А. Дородницына РАН, г. Москва

oli@ultimeta.ru, k.kuznetcov@gmail.com, serebr@ultimeta.ru, vicoff@yandex.ru

Аннотация

Приводится описание портала пространственных данных «ГеоМета» – стандартизированной и децентрализованной среды управления пространственной информацией, разработанной для доступа к базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников. Рассматриваются основные возможности портала, приводятся функции системы поддержки портала и ее архитектура.

1 Предпосылки создания системы

Организации Российской академии наук и смежных ведомств обладают огромным количеством разнообразной пространственной информации, распределенные источники которой отличаются качеством и полнотой представления информации, включают данные справочного характера и картографические данные.

Карты – эффективные инструменты коммуникации – играют важную роль в работе многих пользователей:

- принимающих решения – планировщиков работ, которым необходимы быстрые, надежные и своевременные картографические продукты с дружелюбным интерфейсом как основа для действий, лучшего планирования и мониторинга;
- GIS-экспертов, которым требуется обмениваться непротиворечивой и актуальной географической информацией;
- пространственных аналитиков, нуждающихся в междисциплинарных данных для осуществления предварительного географического анализа и надежных прогнозов при выборе действий в соответствующих областях.

К настоящему времени в учреждениях РАН накоплен большой опыт использования геоинформационных технологий, реализованы многочисленные геоинформационные проекты, созданы базы и банки пространственных данных.

Академические ресурсы пространственных данных составляют значительную часть национальных информационных ресурсов. Основным производи-

телем пространственных данных являются учреждения геологического, геофизического, географического и экологического (природоохранного) профилей. В то же время данные рассредоточены, их использование ограничено зачастую рамками того проекта, где они созданы, затруднены или невозможны поиск существующих данных и доступ к ним, не налажен обмен ими. Причина этого – отсутствие эффективной системы управления пространственными данными. Ее создание позволит интегрировать данные и знания о территории, строить и использовать модели природных и социально-экономических явлений и процессов, их взаимодействия в системе «общество – природная среда», использовать методы пространственного анализа, обеспечивать территориальное планирование и управление.

В целом в учреждениях РАН имеется опыт выполнения разнообразных геоинформационных проектов для различных приложений, сформированы подразделения, отделы и лаборатории геоинформатики, укомплектованные высокопрофессиональными научными кадрами, располагающими необходимой технической базой, современными программными средствами геоинформационных систем (ГИС) и данными, то есть созданы необходимые условия для разработки ГИС и их интеграции.

Основным инструментом интеграции и предоставления пространственных данных в настоящее время являются геопорталы. Понятие «геопортал» означает точку входа в интернет с инструментами просмотра метаданных, поиска географической информации, ее визуализации, загрузки, распространения и, возможно, поиска геосервисов.

Современное требование к системам поддержки геопорталов – независимость, расширяемость и гибкость компонентов, являющаяся важной особенностью современной программной системной архитектуры.

Существует потребность в объединении этих данных, имеющих распределенный характер, в концептуально одну информационную систему, в обеспечении централизованного доступа к ним, в создании на основе интернета технологий единого информационного пространства геоданных.

2 Основные возможности системы

Портал «ГеоМета» – это стандартизированная и децентрализованная среда управления пространственной информацией, разработанная для доступа к

Труды 12^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010

базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников, облегчающая обмен пространственной информацией между организациями и ее совместное использование посредством интернета. Этот подход к управлению географической информацией имеет целью предоставить широкому сообществу пользователей средства для простого и своевременного доступа к имеющимся пространственным данным и существующим тематическим картам, которые могут оказаться полезными для поддержки информированного принятия решений.

Главная цель портала – увеличить доступность разнообразных междисциплинарных данных различного масштаба вместе с сопутствующей информацией, организованных и документированных стандартным и непротиворечивым способом, улучшить кооперацию и координацию усилий при сборе данных, сохраняющих ресурсы и, в то же самое время, ограждающих данные и информацию от нежелательного доступа.

Портал «ГеоМета» представляет собой платформу для создания распределенной среды интеграции неоднородных источников геоинформационных данных и предоставления к этой среде единой точки входа (веб-портала), которая позволит ученым в сфере наук о Земле легко находить специализированные данные и приложения, производить вычислительные эксперименты, визуализировать результаты деятельности.

Благодаря тому, что портал «ГеоМета» построен на базе ИС «НИ РАН» [1], являющейся базовым инфраструктурным компонентом ЕНИП [2], он может интегрироваться в ЕНИП с предоставлением расширения схемы геопропространственными метаданными и геоданными.

К функциональностям ГИС-части системы относятся:

- каталогизация, сбор, поиск геопропространственных метаданных;
- размещение геоданных в собственном хранилище и предоставление к ним доступа;
- предоставление доступа к распределенным геопропространственным данным по стандартизованным протоколам;
- визуализация карт, редактирование элементов.

Интерфейс системы представлен веб-порталом, поэтому для ГИС-части основным методом доступа пользователя к информации является обычный доступ к веб-страницам портала через любой распространенный браузер. Ядро системы предоставляет следующие возможности: управление статическим содержанием; хранение объектов системы (представленных RDF-тройками) в реляционных СУБД; индексирование и полнотекстовый поиск; обеспечение безопасности системы.

Система поддерживает следующие основные типы ресурсов: *Пространственные данные* (картографические данные и их метаданные) и дополнительные типы ресурсов, такие, как *Организация*, *Персона*, *Публикация*, *Проект* и различные рубри-

каторы и классификаторы.

Ресурс *Пространственные данные* содержит наборы пространственных данных и метаданные распределенных пространственных данных. Ресурс *Организация* включает организации РАН, научные центры и другие организации. Данные об их сотрудниках сопоставлены ресурсу *Персона*. Ресурс *Проект* поддерживает сведения о проектах, выполненных или ведущихся в РАН и других ведомствах. Ресурс *Публикация* представляет данные о публикациях и научной деятельности.

Доступ к portalу осуществляется интерактивно через интернет посредством веб-браузера (например, Netscape Navigator или Microsoft Internet Explorer) по ссылке <http://www.geometa.ru>.

В данной работе основной акцент сделан на расширениях, реализованных по сравнению с предыдущей версией портала [3], а именно, на интеграции дополнительных идентификационных данных со средой ЕНИП [1, 2], расширениях поиска, просмотра и регистрации ресурсов и средствах периодического сбора метаданных.

В предшествующей версии портала были представлены средства регистрации, поиска и просмотра метаданных геосервисов. В связи с интеграцией в систему полноценного геосервера все широко распространенные геосервисы фактически имеются и существенно используются в системе. Однако разработчики «ГеоМета» не считают продуктивным предоставлять непосредственный доступ к геосерверу со стороны внешних систем, и поэтому в настоящее время ведется разработка модуля дружественного пользовательского интерфейса использования геосервисов. До окончания этой разработки средства работы с метаданными геосервисов скрыты, хотя и остаются в рабочем состоянии в соответствии с их описанием в [3].

3 Пространственные данные и метаданные

Информационной основой для поиска данных, оценки их пригодности для конкретной задачи и доставки потребителю являются метаданные.

Публикация метаданных позволяет производителям и хранителям пространственных данных сообщать потенциальным потребителям о том, какими ресурсами они располагают и где они хранятся. Соответственно потребители получают возможность вести поиск данных, которые им нужны для решения своих задач, причем метаданные важны не только для внешней публикации, но и для ведения каталогов собственных информационных ресурсов организаций.

При создании геопортала «ГеоМета» была поставлена цель разработать простой, но достаточно полный профиль метаданных [3] для пространственных данных и сервисов, ориентированный на максимальную совместимость, адаптируемость и расширяемость как для российских, так и зарубежных поставщиков и потребителей пространственных ресурсов.

Задача интеграции распределенных неоднородных источников пространственных данных и сервисов решается путем выделения базовой схемы метаданных, реализованной в виде OWL-онтологии [5] на основе стандартов ISO 19115:2003 «Географическая информация. Метаданные» [7] и ISO 19139:2007 «Географическая информация. Метаданные – XML схема» [8] и полностью включающей российский стандарт ГОСТ Р 52573-2006 «Географическая информация. Метаданные» [4] (а также учитывающей профили стандартов и схемы метаданных, лежащие в основе наиболее известных национальных инфраструктур пространственных данных (ИПД)), для обеспечения семантической интероперабельности систем, участвующих в распределенном взаимодействии.

В связи с принятием российского стандарта в системе «ГеоМета» учитывается, но непосредственно не реализован широко распространенный американский стандарт FGDC (Federal Geographic Data Committee (<http://www.fgdc.gov>)). Стандарты OpenGIS [9] использовались при разработке архитектуры, интерфейсов и протоколов обмена системы.

Элементы метаданных профиля «ГеоМета» поддерживают следующие функции работы с метаданными:

- поиск информации, необходимой для определения наборов данных, имеющих на определенную географическую область;
- определение назначения и пригодности информации (оценка соответствия набора данных определенным потребностям);
- осуществление доступа к информационным ресурсам, которые необходимы для получения выбранного набора данных и сервисов;
- применение ресурсов, т. е. обработка и использование наборов данных и сервисов.

Для моделирования схемы и структуры метаданных в стандартах ISO 191XX и их профилях используется язык UML.

С учетом того, что в России принят национальный стандарт ГОСТ 52573-2006, именно на его основе используется соответствующий технический регламент, учитывающий все вышеизложенные функции метаданных.

Кроме того, в этом же регламенте определена система уникальной идентификации наборов пространственных данных, которая необходима для обмена этими данными.

Согласно ISO 19115, метаданные геоинформационного ресурса делятся на группы, называемые пакетами. Каждый пакет содержит информацию об определенном аспекте описываемого набора пространственных данных (кроме первого пакета, который относится к самим метаданным): характеристики набора метаданных; идентификация; ограничения на доступ и использование; качество данных (согласно ISO 19113); информация об обновлениях; организация пространственных данных в наборе; система координат; содержание; применимый набор условных знаков; возможность приобретения; пользовательские расширения метаданных; схема при-

ложения; пространственно-временная протяженность; ответственная сторона и правила цитирования.

Пакет состоит из одной или более сущностей, связанных отношениями обобщения (агрегирования). При необходимости сущности могут повторяться в рамках одного пакета для решения задач пользователя. Каждая сущность представляет собой совокупность элементов (атрибутов), характеризующих тот или иной аспект метаданных. Сущности или элементы метаданных могут быть обязательными, необязательными или условными.

Данные в пакетах сгруппированы в разделы, например, в пакете Идентификационная информация о данных имеются разделы: Общие сведения, Идентификационные данные, Пример изображения для просмотра, Ключевые слова и Периодичность обновления.

4 Интеграция дополнительных идентификационных данных

В стандартах ISO информация об Организациях и Персонах представлена отдельными классами, но при вводе и выдаче этих данных в большинстве ГИС наблюдаются дублирование с повышением вероятности ошибок и несовпадения значений атрибутов. Кроме того, наборы атрибутов по указанным стандартам весьма ограничены по сравнению с наборами соответствующих объектов ИС «НИ РАН» (а в стандартах поощряется расширение дополнительных объектов в прикладных профилях метаданных). В частности, атрибутика объектов «НИ РАН» соответствует рекомендациям стандарта Dublin Core (<http://dublincore.org/>) и других международных стандартов на метаданные общего характера.

В связи с этим в портале «ГеоМета» проведена интеграция дополнительных идентификационных данных среды ЕНИП (для персоналии и организаций) в средства регистрации, хранения и визуализации пространственных метаданных с сохранением возможностей просмотра и редактирования этих данных средствами среды ЕНИП, а также выборки метаданных, связанных с данными, найденными средствами ЕНИП.

Это позволяет обеспечить важнейший принцип организации и функционирования системы – взаимосвязь ресурсов. В частности, информация об ответственном лице некоторого пространственного ресурса связана с ресурсом «Персона», который средствами ЕНИП связан с «Организациями», а список публикаций связан как с ресурсами «Персона» (авторы публикации), так и с ресурсом «Публикация» (собственно публикация).

Прежде всего, в OWL-онтологию и в схему базы данных хранилища системы были введены элементы обеспечения связи объектов пространственных данных со стандартными объектами персоналии и организаций в ЕНИП (и, соответственно, удалены старые дублирующие классы в пространственных метаданных). Соответствующим образом скорректированы функции поддержки целостности при мо-

дификации и удалении взаимосвязанных данных.

Пользовательский интерфейс ввода метаданных обеспечен элементами и страницами связи со стандартными функциональными модулями поддержки объектов персоналии и организаций в ЕНИП. Значения полей этих объектов теперь могут быть выбраны из раскрываемых списков.

При отсутствии требуемого дополнительного объекта происходит переход к стандартным средствам ЕНИП по вводу или модификации значений этого объекта. Аналогично в средствах визуализации метаданных информация о дополнительных объектах сопровождается ссылкой на стандартные страницы выдачи подробной информации об этих объектах. В то же время, в страницах выдачи метаданных в формате XML обеспечено дублирование значений дополнительных объектов в соответствии со стандартом 19139:2007.

5 Поиск основных ресурсов

При нажатии пункта меню «Метаданные» в разделе «Расширенный поиск» основного меню будет выдана страница поиска метаданных ресурсов.

The screenshot shows a web interface for searching metadata resources. On the left, there are several input fields: 'Название:' (Name), 'Предметная область:' (Subject area) with a dropdown menu, and 'Ключевые слова:' (Keywords). Below these is a section for 'Тип метаданных:' (Metadata type) with radio buttons for 'Ресурс' (Resource) and 'Слой' (Layer). To the right is a map titled 'Выбор координат на карте:' (Select coordinates on the map) with navigation controls. Below the map is a 'Ввод координат:' (Enter coordinates) section with input fields for 'Север' (North) (71.71875), 'Запад' (West) (-127.6171875), and 'Восток' (East) (81.2109375), along with a 'Очистить' (Clear) button and a 'Юг' (South) field with the value -68.90625. At the bottom are 'Поиск' (Search) and 'Сбросить' (Reset) buttons.

Рис. Страница поиска основных ресурсов

На этой странице можно заполнить нужные поисковые поля, указать объекты поиска – слои или полные ресурсы, ввести или выбрать курсором на карте географические координаты для искомых ресурсов. При нажатии кнопки «Поиск» выполняется поиск ресурсов и результат выдается в формате «Каталог ресурсов».

Можно также войти через пункт меню «Тематические разделы» в разделе «ГИС» – будет выдан список тематических разделов (по стандарту ISO) с указанием наличия для каждого пункта количества ресурсов с данной тематикой.

При нажатии на некоторый раздел выдается список его ресурсов в формате «Каталог ресурсов».

Для просмотра основных пространственных ресурсов можно выполнить их поиск или нажать пункт меню «Каталог ресурсов» в разделе «ГИС» и перейти на страницу Каталога ресурсов.

Через пункт меню «Информация о свободных слоях» можно выдать список указателей на метаданные слоев, не присоединенных к какому-либо ресурсу.

6 Просмотр метаданных и пространственных данных

Нажатие на название ресурса производит переход на страницу показа его метаданных, при нажатии на ссылку «Просмотреть в ГИС-клиенте» переходим к просмотру карты ресурса. Для обеспечения показа карт в составе системы имеются две заимствованные компоненты: средство для хранения и публикации геоданных (геосервер) и клиентское средство визуализации геоданных (ГИС-клиент) Openlayers.

На странице просмотра метаданных в верхней левой части содержатся ссылки на страницы выдачи метаданных в формате XML и выдачи метаданных о слоях ресурса с указанием количества слоев.

Далее выдаются метаданные ресурса в формате профиля системы «GeoMeta», соответствующем стандарту ISO 19115:2003. Для каждого свойства метаданных выдаются его английское и русское наименование, значение и номер в справочнике ISO 19115:2003. Сведения об ответственных лицах и организациях в метаданных выдаются в виде ссылок на более полные записи в дополнительных ресурсах системы.

Метаданные слоев в основном аналогичны метаданным ресурсов, но имеются отличия, в частности, как правило, присутствует ссылка на изображение.

При нажатии ссылки «Просмотр в XML» выдается страница текста метаданных в формате ISO 19119. Эти данные могут быть скопированы и загружены в любом портале, поддерживающем загрузку метаданных в данном стандарте.

Просмотр пространственных данных осуществляется обращением к ГИС-клиенту, реализованному на основе Openlayers.

ГИС-клиент содержит следующие элементы: кнопки управления смещением карты; кнопки управления масштабом; режим управления картой мышью; режим выбора области просмотра; панель выбора слоев; мини-карта; масштабная линейка; панель отображения легенды карты.

При одиночном клике мышью на карте выводится всплывающее окно с результатами соответствующего GetFeatureInfo-запроса к WMS-серверу. При этом WMS-сервер со своим хранилищем геоданных может находиться на каком-либо внешнем портале или же могут использоваться WMS-сервер и его база данных, расположенные непосредственно в портале «GeoMeta». Среди ресурсов портала «GeoMeta» имеются примеры обоих вариантов.

7 Регистрация пространственных ресурсов

При входе в регистрацию пространственных ресурсов пользователь попадает на первую страницу ввода основных идентификационных данных о ресурсе. В верхней части этой и последующих страниц регистрации метаданных расположено меню четырех пакетов метаданных профиля «GeoMeta»: Идентификационная информация о данных, Инфор-

мация о координатной основе, Информация о распространении и Информация о метаданных. При нажатии на какой-либо из пунктов этого меню под ним появляется меню разделов данного пакета (если они имеются). При нажатии на один из его пунктов выдается страница ввода данных этого раздела.

При регистрации слоя на вкладке «Загрузка SHP» можно осуществить загрузку shape-файла слоя в хранилище геоданных портала «ГеоМета». Shape-файл должен быть запакован в zip-архив, содержащий .shp файл и, возможно, некоторые другие файлы (.prj, .dbf и т. п.). При этом из файла будут автоматически выбраны и сохранены в системе некоторые метаданные, а именно, название слоя, EPSG-код проекции и ограничивающий прямоугольник.

В разделе «Регистрация» основного меню имеется вход «Загрузка метаданных в стандарте ISO 19139». Здесь можно указать и загрузить файл или zip-архив с файлами в формате стандарта ISO 19139 из директории пользователя или по указанному URL.

Поиск, просмотр и регистрация дополнительных ресурсов выполнены стандартными средствами ИС «НИ РАН» и подробно описаны в документации и публикациях по этой системе [1].

8 Периодический сбор метаданных (harvesting)

В портале «ГеоМета» предусмотрена возможность периодического сбора метаданных (harvesting) из различных источников – каталогов метаданных. При этом узел «ГеоМета» может выступать как в роли клиента, который выполняет сбор информации из других каталогов, так и в роли сервера, который может предоставлять метаданные в другие каталоги.

Каждый каталог, участвующий в сборе, может предоставлять в другие каталоги как локальные метаданные, так и метаданные, собранные в нем с удаленных источников. Механизм периодического сбора метаданных обеспечивает копирование метаданных с удаленного источника и их синхронизацию с заданной периодичностью. Управляет сбором метаданных Администратор системы, который определяет список источников каталогов, участвующих в сборе, и периодичность синхронизации метаданных.

В настоящее время реализованы два протокола обмена: GeoNetwork – порталы на базе ресурсов системы «Geonetwork opensource» [6] и Web DAV – каталоги на базе сервера WebDAV, используются для обмена с другими порталами, кроме GeoNetwork, и, в частности, для обмена между системами «ГеоМета». В обоих случаях текст метаданных передается в формате XML по стандарту ISO 19139:2007.

Обмен метаданными между системами «ГеоМета» и «GeoNetwork» выполняется по стандартному протоколу «GeoNetwork 2.1» для сбора данных с узлов типа «GeoNetwork». Для «GeoNetwork» мета-

данные с узлов «ГеоМета» доступны по протоколу WebDav.

В портале имеется Страница списка узлов опроса с указанием свойств каждого узла: Название, URL, Тип узла и Период опроса. Сбор данных запускается вручную Администратором или происходит периодически, в соответствии с указанным для узла периодом.

При нажатии на элемент управления «Добавить новый узел» выдается страница регистрации узла.

Протокол обмена «GeoNetwork» подробно описан в документации по этой системе [6]. Протокол обмена «ГеоМета» основан на использовании сервера WebDAV, позволяющего считывать файлы из указанной директории удаленного источника. В этой директории для успешного выполнения сбора должны содержаться файлы метаданных портала в формате ISO 19139:2007. Все файлы данного формата в указанной директории будут в момент сбора выбраны для перерегистрации метаданных.

9 Заключение

Портал «ГеоМета» является результатом научно-исследовательской работы «Создание пилотного сегмента распределенной среды геопространственных данных и вычислительных приложений», целью которой является создание академической инфраструктуры пространственных данных (АИПД) для обеспечения хранения и использования распределенных геопространственных данных и вычислительных приложений в целях организации единого информационного пространства научных исследований в области наук о Земле. Используются ключевые национальные, международные и отечественные стандарты, применяемые при разработке АИПД, ее общая архитектура и основные элементы. Обосновано применение базовых принципов и технологий Semantic Web, использование онтологий и форматов XML/RDF/RDFS/OWL для описания компонентов распределенной среды данных и приложений.

В проекте реализованы средства интеграции приложений с разными источниками/каталогами метаданных/данных, сервис директорий ГИС метаданных (OWL схема ISO 19115 на основе схем ЕНИП, формат обмена на основе ISO 19139), унифицированный интерфейс поиска геопространственных данных, подсистема взаимодействия портала с локальным ГИС-сервером, подготовлены и размещены на геопортале метаданные базовых наборов пространственных данных.

Особое внимание уделено стандартам геометаданных и методам реализации платформы для создания распределенной среды интеграции неоднородных источников геоинформационных данных, а также практике использования АИПД в некоторых работах, выполняемых совместно с организациями РАН, тесно связанными с геопространственными данными.

Предлагаемые базовые средства и механизмы распределенной среды геопространственных дан-

ных и приложений позволят ученым перейти на новую ступень обработки и компоновки данных и приложений, которая в настоящий момент получила название «электронная наука» (e-science). Пройдя апробацию в научной среде, вновь созданные компоненты могут быть выведены на рынок программного обеспечения в качестве типовых средств интеграции распределенной гетерогенной среды геоданных и приложений (сервисов).

Литература

- [1] Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М., Теймуразов К.Б., Филиппов В.И. Информационная Web-система «Научный институт» на платформе ЕНИИП. – М.: ВЦ РАН, 2007. – 254 с.
- [2] Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН. – М.: Вычислительный центр РАН, 2006. – 238 с.
- [3] Бездушный А.Н., Вершинин А.В., Дьяконов И.А., Динь Ле Дат, Серебряков В.А. Пространственные метаданные в системе «GeoMeta» // Журнал «Пространственные данные». – 2008. – № 2, 3.
- [4] ГОСТ Р 52573–2006 – Географическая информация. Метаданные. – <http://protect.gost.ru/document.aspx?control=7&id=74833>.
- [5] Динь Ле Дат, Серебряков В.А. Разработка и реализация формальных онтологий геопространственных данных и сервисов // Журнал «Радиотехника». – 2008. – Вып. 2 – С. 85-89.
- [6] Geonetwork opensource. The complete manual. V 2.4, 2009. – <http://geonetwork-opensource.org/>.
- [7] ISO 19115:2003 – Geographic information. Metadata.
- [8] ISO 19139:2007 – Geographic information – Metadata – XML schema implementation.
- [9] OpenGIS Specifications (Standards). – <http://www.opengeospatial.org/standards/>.

Spatial data integration portal "GeoMeta"

O.M. Ataeva, K.A. Kuznetsov, V.I. Filippov,
V.A. Serebriakov

The spatial data portal GeoMeta is described as a standardized and decentralized spatial information management environment, designed for access to geodata, cartographic products and related metadata from various sources. The main features of the portal are discussed, the functions of the portal support system and its architecture.

* Работа выполнена при финансовой поддержке РФФИ (проект 09-07-12065-офи_м)