

Среда интеграции пространственных данных «ГеоМета»*

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

Вычислительный центр РАН

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

Аннотация

Приводится описание стандартизированной и децентрализованной среды управления пространственной информацией на основе портала пространственных данных «ГеоМета», разработанного для доступа к базам геоданных, картографическим продуктам и связанным с ними метаданным из различных источников. Рассматриваются основные возможности портала с акцентом на расширения, реализованные в его последних версиях.

1. Основные возможности «ГеоМета»

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

Основная 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 уче-

ным 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 геопространственными метаданными и геоданными.

К функциональностям ГИС-части системы относятся:

- каталогизация, сбор, поиск геопространственных метаданных;
- размещение геоданных в собственном хранилище и предоставление к ним доступа;
- предоставление доступа к распределенным геопространственным 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 рубрики 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. Реализованы дополнительные по сравнению со стандартом ISO средства связи ресурсов *Проект* и *Публикация* с пространственными метаданными.

Труды XIV Всероссийской объединенной конференции «Интернет и современное общество» (IMS-2011), Санкт-Петербург, Россия, 2011.

Доступ к portalу осуществляется интерактивно через интернет посредством любого 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 предыдущими версиями portalа [3,5], а именно, на интеграции дополнительных идентификационных данных со средой ЕНИП [1, 2], расширениях поиска, просмотра и регистрации ресурсов и средствах периодического сбора метаданных.

2. Геометаданные и пространственные данные

Информационной основой для поиска геоданных и оценки их пригодности для конкретной задачи и предоставления пользователю 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 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 определенной географической области;
- 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 приложения; 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.

3. 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 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.

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).

В связи с этим в OWL-онтологию и в схему базы данных хранилища системы были введены элементы обеспечения связи объектов пространственных данных со стандартными объектами персоналий, организаций, публикаций и проектов в ЕНИП (и, соответственно, удалены старые дублирующие классы в пространственных метаданных). Соответствующим образом скорректированы функции поддержки целостности при модификации 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.

4. Поиск пространственных ресурсов

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

На этой странице можно заполнить нужные поисковые поля, 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.

5. 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 (метаданные о слоях даются только для ресурсов, геоданные которых содержатся в данном экземпляре портала).

Далее выдаются метаданные ресурса в формате профиля системы «ГеоМета», 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 для внутренних ресурсов портала 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 19139. 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 выводит-ся 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 «ГеоМета». Среди ресурсов портала «ГеоМета» имеются примеры обоих вариантов.

Однако, внешняя ссылка на георесурс может представлять собой не только указание на WMS-сервер, но и просто Интернет-ссылку на какой-либо сайт или его элемент, – в этом случае геосервер и ГИС-клиент не используются, а происходит обращение по указанной ссылке.

6. 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 «ГеоМета»: Идентификационная информация о данных, Информация о координатной основе, Информация о пространстве 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].

7. Дополнительные средства администрирования

В последней версии системы реализована подсистема управления доступа к ресурсам для пользователей. Ресурсы подразделяются на две категории

проверенные и не проверенные. В общий каталог доступный для просмотра как авторизованным так и не авторизованным пользователям попадают только ресурсы которые имеют статус «проверенный». Этот статус может менять только администратор системы. Это нововведение было связано с тем, что как оказалось во время харвестинга настроенного не точно в систему попадает большое число «мусорных» ресурсов, которые засоряют каталог. Теперь каталог содержит только проверенные и одобренные администратором системы ресурсы. Для авторизованного пользователя в системе появилась возможность регистрации всех видов ресурсов. Доступ же ко всем действиям над ресурсами возможен только в том случае, если конкретный ресурс создавался самим пользователем, иначе пользователь может только просматривать ресурс. Также авторизованный пользователь получил возможность просматривать созданные им ресурсы из пункта меню "Каталог пользователя", которые имеют статус «не проверенный». Для того чтобы пользователь имел доступ к описанным возможностям, он должен зарегистрироваться в системе и администратор должен включить его в специальную группу с соответствующими правами доступа к системе. Для администраторов системы появилась возможность отслеживать все ресурсы в системе через пункт в меню «каталог администратора».

Теперь при просмотре данных в системе на странице просмотра в правом нижнем углу всегда отображается информация о том, кто из пользователей создал этот ресурс и когда.

У администратора портала появилась возможность завести рубрикатор, с помощью которого можно классифицировать ресурсы не по стандарту ISO, а в соответствии с интересами поддерживающей портал организации. Каждый ресурс может принадлежать нескольким элементам рубрикатора. При необходимости можно ввести несколько рубрикаторов. При просмотре рубрикатора также имеется возможность осуществить поиск ресурсов по узлам этого рубрикатора.

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» метаданные с узлов «ГеоМета» доступны по протоколу WebDav.

Механизм сбора метаданных основан на концепции использования универсальных уникальных идентификаторов метаданных. Это специальный id, который уникален не только в пределах конкретной системы, но и за ее пределами. Он представляет собой комбинацию текущей даты, MAC-адреса и случайного числа.

Дата последнего изменения метаданных в системе также играет важную роль, позволяя принимать решения о необходимости повторного сбора метаданных и обновления их в принимающей системе.

Опираясь на универсальные уникальные идентификаторы и дату последнего изменения при выполнении сбора метаданных, принимаются решения о необходимости добавления, обновления, удаления метаданных. Таким образом при сборе метаданных не происходит их дублирования.

Если одни и те же метаданные пришли из различных источников, благодаря использованию универсальных уникальных идентификаторов и даты последнего изменения имеющиеся данные будут корректно изменены.

При нажатии на элемент управления «Добавить новый узел» выдается страница регистрации узла. Здесь следует указать «Название» узла (уникальное), URL узла (в соответствии с протоколом), «Тип узла» – из выпадающего списка выбирается наименование используемого протокола обмена: «ГеоМета» или GeoNetwork, – и период автоматического запуска опроса для данного узла (в днях). Еще можно добавить критерий отбора (поиска) метаданных для данного узла, указав поиск заданного текста во всех текстовых полях метаданных или в одном или комбинации полей «Название», «Ключевые слова» и «Краткое содержание».

Рис. 1. Страница узла харвестинга

После создания узла харвестинга сбор данных осуществляется вручную администратором при нажатии на соответствующую кнопку. При этом запускается сбор данных и метаданные сохраняются в системе. Так как сбор данных затем происходит периодически, то любые изменения метаданных в локальной системе будут утеряны при следующем повторе, и поэтому данные, которые попали в систему таким образом, недоступны для редактирования.

Процесс сбора метаданных может прерываться в следующих случаях:

1. Остановка администратором.
2. Возникновение ошибки.

При удалении узла сбора метаданных все принадлежащие ему ресурсы могут быть удалены из системы по решению пользователя.

В портале имеется Страница списка узлов опроса с указанием свойств каждого узла: Название, URL, Тип узла и Период опроса.

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 отечествен-

ные стандарты, применяемые при разработке АИПД, ее общая архитектура и основные элементы.

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 геосервером, подготовлены и размещены на геоportale 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.

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. Информационная Web-система «Научный институт» на платформе ЕНИП / Бездушный А.А., Бездушный А.Н., Нестеренко А.К., Серебряков В.А., Сысоев Т.М., Теймуразов К.Б., Филиппов В.И. – М.: Вычислительный центр РАН, 2007.
- [2] Бездушный, А.А. Интеграция метаданных Единого Научного Информационного Пространства РАН / Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. – М.: Вычислительный центр РАН, 2006. – 238 с.
- [3] Бездушный, А.Н. Пространственные метаданные в системе «ГеоМета» / Бездушный А.Н., Вершинин А.В., Дьяконов И.А., Динь Ле Дат, Серебряков В.А. // Пространственные данные. – 2008. – № 2. № 3.
- [4] ГОСТ Р 52573–2006 – Географическая информация. Метаданные. [Электронный ресурс]. — Режим доступа: <http://protect.gost.ru/document.aspx?control=7&id=74833>.
- [5] Атаева О.М., Кузнецов К.А., Серебряков В.А., Филиппов В.И. Портал интеграции пространственных данных «ГеоМета». – М.: ВЦ РАН, 2010. – 106 с.
- [6] Geonetwork opensource. The complete manual. V 2.4, 2009. [Электронный ресурс]. — Режим доступа: <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). [Электронный ресурс]. — Режим доступа: <http://www.opengeospatial.org/standards/>.

Spatial data integration environment «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, with the accent on the expansions, realized in its last versions.

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