

**МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. Ломоносова
Факультет Вычислительной Математики и Кибернетики**

На правах рукописи



Динь Ле Дат

**РАЗРАБОТКА И РЕАЛИЗАЦИЯ ФОРМАЛЬНЫХ
ОНТОЛОГИЙ ПРОСТРАНСТВЕННЫХ
ДАнных И СЕРВИСОВ**

Специальность 05.13.11 -
математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2008

Работа выполнена на кафедре Системного программирования факультета вычислительной математики и кибернетики Московского Государственного Университета им. М.В.Ломоносова

Научный руководитель:

доктор физ.-мат. наук, профессор
Серебряков Владимир Алексеевич

Официальные оппоненты:

доктор физ.-мат. наук, профессор
Соловьев Сергей Юрьевич

кандидат географических наук
Кошкарев Александр Владимирович

Ведущая организация:

Дальневосточный геологический институт
Дальневосточного отделения РАН

Защита диссертации состоится «04» апреля 2008 года в 11:00 час. на заседании диссертационного совета Д 501.001.44 в Московском государственном университете им. М.В. Ломоносова по адресу: 119992, ГСП-2, Москва, Ленинские горы, МГУ, 2-й учебный корпус, факультет ВМК, аудитория 685.

С диссертацией можно ознакомиться в библиотеке факультета ВМК МГУ. С текстом автореферата можно ознакомиться на официальном сайте ВМиК МГУ им. М.В. Ломоносова <http://www.cmc.msu.ru> в разделе «Наука» - «Работа диссертационных советов» - «Д 501.001.44»

Автореферат разослан «04» марта 2008 г.

Ученый секретарь
диссертационного совета,
профессор



Н. П. Трифонов

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы

Эффективность использования геоинформационных систем (ГИС) и технологий в различных сферах трудовой деятельности человека определяется прежде всего тем фактом, что более 80% информации, с которой сталкивается человек в своей жизни, имеет территориальную привязку. Поэтому на сегодняшний день не вызывает сомнения потребность в разработке многопользовательских многофункциональных Интернет-ГИС-приложений, которые, в свою очередь, ставят задачу совершенствования существующих и разработку новых подходов к сбору, хранению, обработке и распространению растущего объема пространственной информации и сервисов. Более того, ГИС в последние несколько лет совершили значительное продвижение в направлениях распределенной работы, совместного использования ресурсов и пр. При этом ввиду интеграции большого количества независимых информационных систем существенной проблемой является различная трактовка структуры и семантики информации в разных системах. Данные могут относиться к различным предметным областям, в рамках одной иметь разные выражение и интерпретацию.

Для того, чтобы управлять процессами создания, хранения, обновления и обработки колоссального объема пространственных данных и услуг, необходимо формировать метаданные. Однако достижения согласия по поводу одной единственной спецификации метаданных в мировом масштабе практически невозможно. Несмотря на рекомендации ISO/TC211, FGDC и OGC на использование стандартов пространственных метаданных, на рынке пространственных данных и сервисов сложилось так, что почти в каждой стране создаются свои профили метаданных. В итоге разные геоинформационные системы и каталоги поддерживают и рекомендуют свои профили метаданных. В частности, с 1 января 2007г. в России был введен в действие стандарт пространственных метаданных ГОСТ Р 52573–2006 «Географическая информация. Метаданные», который является профилем ISO 19115 – международного стандарта пространственных цифровых метаданных. К сожалению, Российский стандарт не имеет полной совместимости с исходным международным стандартом, поэтому его необходимо откорректировать, чтобы не создавать новых технических барьеров. В связи с этим, разработка или адаптация спецификации метаданных, которая бы помогла описать формат и содержание данных для достижения интероперабельности, является актуальной задачей нескольких инициативных групп, представляющих различные науки о Земле.

Наиболее эффективным способом обеспечения согласованного информационного обмена в таких условиях является обеспечение так называемой семантической интероперабельности систем на основе стека технологий семантического веба (Semantic Web): RDF, RDFS, OWL. Язык Web-онтологий OWL представляет развитые возможности для описания предметных областей взаимодействующих систем и дает машинно-интерпретируемые определения фундаментальных понятий в предметной области и отношениях между таковыми в онтологии.

В рамках данного исследования анализируются международные и российские спецификации пространственных метаданных, лежащих в основе наиболее известных в мире инфраструктур пространственных данных (ИПД) и предлагается простой, но в то же время достаточно полный профиль метаданных для пространственных данных и сервисов, ориентированный на максимальную совместимость, адаптируемость и расширяемость как для российских, так и для зарубежных поставщиков и потребителей геоинформационных ресурсов, что и привело к онтологическому подходу формирования метаданных.

С другой стороны, основной целью проводимых исследований является создание в рамках Единого научного информационного пространства (ЕНИП) РАН геоинформационно-аналитической системы Web-портала (гео-портал) «ГеоМЕТА», основанной на интеграции децентрализованно развиваемых пространственных ресурсов институтами, входящими в РАН, и запуск в работу онлайн-механизмов автоматического обмена пространственными метаданными между информационной системой РАН на базе ЕНИП РАН. Следовательно, решение проблемы повышения качества программного обеспечения ГИС в рамках ЕНИП также обуславливает актуальность решаемой в диссертации научной задачи.

Цель работы

Цель работы заключается в исследовании и разработке геоинформационно-аналитической системы интеграции распределенных неоднородных источников пространственных данных и сервисов на основе пространственных онтологий и решении проблемы семантической интероперабельности, а также задачи управления метаданными и совершенствования поиска, доступа и обмена в условиях растущих объемов пространственной информации и сервисов, предоставляемых многочисленными источниками геоинформации.

В соответствии с целью работы были поставлены следующие задачи:

1. Разработать онтологию пространственных метаданных для каталогизации, сбора и поиска геоданных и сервисов совместимо с ISO 19115:2003, ГОСТ Р 52573–2006 и ISO 19119:2005 «Geographic information -- Services».

2. Разработать онтологию, основанную на семантике существующих стандартов контента, таксономиях и ключевых словах в области наук о Земле.
3. Расширить возможности описания геосервисов (метаданных услуг) за счет использования онтологии для придания значения контенту данных, улучшив тем самым семантическую функциональную совместимость.
4. Создание и внедрение на практике геопортала и интеграции Web-ресурсов на основе разработанных онтологий для нахождения и обеспечения простого доступа к имеющимся распределенным пространственным данным и сервисам по стандартизованным протоколам (OGC, ISO/TC211 и W3C).

Объект исследования

Объектом исследования являются проблема достижения интероперабельности, каталогизации и сбора метаданных пространственных ресурсов и технологии построения информационно-аналитических геопространственных Web-порталов.

Предмет исследования

Предметом исследования являются подходы и методы использования семантических технологий в геопространственных Web-порталах для реализации информационных процессов на семантическом уровне.

Методы исследования

В ходе диссертационного исследования были использованы модели и методы, основанные на методологии представления и обработки пространственно-распределенной информации в геоинформационных Интернет-системах, теории метаданных и теории моделей, объектно-ориентированных методах построения и программирования программных систем.

Научная новизна

Научная новизна работы заключается в следующем:

1. Предложен новый подход к процессу разработки геоинформационных систем. Новизна подхода, прежде всего, заключается в использовании онтологии пространственных метаданных. В отличие от существующих методик, такой подход обеспечивает:
 - а. описание свойств классов объектов предметной области на языке веб-онтологий OWL, дающий машинно-интерпретируемые определения фундаментальных понятий в предметной области и отношениях между таковыми в онтологии, а не на языке программирования;

- b. уникальное представление информационных объектов, создаваемых на основе описаний URI, для всех подсистем ГИС, что значительно упрощает построения распределенных геоинформационных систем;
 - c. упрощение процесса сопровождения ГИС, поскольку снижается вероятность необходимости изменения программного кода и структур баз данных при внесении изменений в иерархию классов предметной области.
2. Предложена онтологическая модель данных для описания и организации многоуровневого хранилища, распределенных наборов пространственных данных, тематических слоев карт и геосервисов с развитыми средствами регистрирования, поиска и доступа к данным и сервисам, использующая предложенную автором онтологию пространственных метаданных. Её отличительными особенностями являются:
- совместимость с международным стандартом ISO 19115:2003 и одновременно полная совместимость с российским стандартом ГОСТ Р 52573–2006;
 - совместимость с популярными международными стандартами, такими как стандарт FGDC-STD-001-1998 в Американской и Канадской ИПД, CEN prEN 12657 в Европейской INSPIRE и Французский ИПД, UK GEMINI v1.0 в ИПД Великобритании, ANZLIC profile в Австралийской и Новозеландской ИПД, ESRI Profile of CSDGM – популярный коммерческий и т.д.;
 - расширение метаданных для описания OGC-совместимых сервисов на основе ISO 19119:2005 и также расширение семантики описания ресурсов на основе разработанных онтологий геоданных и сервисов;
 - поддержки двуязычного (русский, английский) словаря терминов, семантически полного и удобного для совместного использования, так как он основан на онтологиях предметной области, которые описаны на OWL;
 - поддержки интероперабельности модели метаданных в XML-формате, соответствующей международному стандарту ISO 19139:2007 «Geographic information -- Metadata -- XML schema implementation».
3. Разработан комплекс программ геоинформационно-аналитического Web-портала, реализующего функции сбора, поиска, категоризации и управления пространственными метаданными, где основное внимание уделено на поддержке интероперабельность спецификаций ГОСТ Р 52573–2006 и ISO 19115:2003 с возможностью описания семантики контента объектов и предоставления рекомендаций с учетом семантики их контента.

Практическая ценность

Построенная формальная онтологическая модель пространственных данных и сервисов представляет собой каноническую модель, позволяющую интегрировать распределенные неоднородные источники геоинформационных ресурсов.

Представленная автором онтология пространственных метаданных позволяет решать проблему «взаимопонимания» (т.е. семантической интероперабельности) геоинформационных систем, участвующих в распределенном взаимодействии. Более того, данная онтология нацелена на поддержку интероперабельности между стандартами ГОСТ Р 52573–2006 и ISO 19115:2003, а также расширение для каталогизации геосервисов основе ISO 19119:2005 и обмена метаданными в структурированном XML-формате по стандарту ISO 19139:2007.

Разработанный геопортал на основе онтологии пространственных данных позволяет оптимизировать сложные процессы формирования и управления пространственными метаданными через веб-интерфейс, обеспечивающий многоуровневую детализацию и поиск пространственных ресурсов на семантическом уровне. Кроме того, геопортал обеспечивает двуязычные словари терминов (русский и английский) и ГИС-визуализацию. При этом созданные пространственные онтологии и геопортал имеют возможности использования в рамках ЕНИП РАН и системах управления знаниями.

Реализованные программные средства имеют широкий круг применения в области интеграции и поиска распределенных неоднородных источников пространственных данных и приложений для решения научных задач исследования Земли и каталогизации и автоматизации сбора распределенных пространственных метаданных на основе OGC-совместимых сервисов.

Апробация работы

Основные результаты работы докладывались и обсуждались на следующих научных конференциях и семинарах:

- Наука & Сотрудничество между Вьетнамом и Россией. VIII и IX международная научная конференция (Москва, 2006-2007).
- XII международная GISnet конференция и выставка Геоинформатики (г. Хошимин, Вьетнам, 2006).
- Научно-практический семинар отдела Геоинформационных систем Государственного геологического музея им. В.И. Вернадского РАН "Проблемы и решения задач в области наук о Земле в распределенной ИНТЕРНЕТ среде" (Москва, 2007).
- Научные семинары лаборатории сетевых информационных технологий Геофизического центра РАН (Москва, 2007).

- Международная конференция ECAI Congress of Cultural Atlases III Time & Space in Eurasia – SCI 2004 (Москва, 2007).
- Научно-практический семинар отдела картографии и аэрокосмических методов Московского филиала Русского географического общества "Пространственные метаданные и геопорталы как средства интеграции геоинформационных ресурсов и сервисов" (Москва, 2008).
- Научные семинары отдела систем математического обеспечения Вычислительного Центра им. А.А. Дородницына РАН (Москва, 2006-2008).

Предложены и апробированы на практике подходы и программно-технологические решения для создания интерактивных распределенных геоинформационных Интернет-систем интеграции главного геопортала Вычислительного центра РАН (<http://earth.ras.ru>) и геопортала Государственного геологического музея им. В.И. Вернадского РАН (<http://earth.jssc.ru>).

Публикации

По теме диссертации опубликовано 4 работ, в том числе 1 из списка изданий, рекомендованного ВАК РФ.

Структура и объем работы

Диссертация состоит из введения, 4 глав, заключения, списка литературных источников из 167 наименований и 4 приложений. Содержит 40 рисунков и 6 таблиц. Работа изложена на 151 страницах.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Введение

Во введении дана общая характеристика диссертации, включая обоснование актуальности темы, объект и предмет исследования, цель и основные результаты работы, ее научную новизну, апробацию работы, список публикаций, структуру и объем диссертации.

Глава I. Геоинформационные системы и современные технологии, проблемы поиска и интеграции распределенных пространственных ресурсов

В первой главе приводится анализ возможностей, проблематики и необходимости современных геоинформационных систем (ГИС).

В настоящее время основной особенностью наук о Земле является пространственная детерминированность и огромные, постоянно увеличивающиеся объемы данных и сервисы, в основе которых лежат географические

объекты различного масштаба от локального геологического тела до Земного шара в целом. Задачей информатики при этом становится синтез и интегральный анализ большого объема многоаспектных пространственных данных и сопоставление результатов исследований для выявления общих закономерностей геопроцессов в пространственно-временных координатах на локальном, региональном и глобальном уровнях. Для решения поставленных задач необходимо привлекать достижения современных геоинформационных технологий – распределенные ГИС, представляющие собой распределенные разнородные базы данных, распределенные вычисления, и стандарты взаимодействия открытых систем, а также создание инфраструктуры пространственных данных (ИПД).

Процесс интеграции пространственных данных и сервисов ставит множество проблем, вызванных, в частности, автономностью и разнородностью схем метаданных и источников, количественными и качественными требованиями к обработке запросов.

Далее в главе рассматриваются открытые международные стандарты обмена пространственными ресурсами и отмечается важность серии стандартов ISO/TC211 19000, ГОСТ Р и спецификации передачи данных консорциума OGC (Open Geospatial Consortium). Особое внимание уделяется изучению наиболее известных открытых географических стандартов, таких как ISO 19119:2005 «Geographic information -- Services», спецификация интерфейса OGC Web Map Service «Интерфейс картографического веб-сервера», спецификация интерфейса OGC Web Feature Service «Веб-сервер объектов» и операции GetCapabilities, позволяющей извлечь метаданные о ресурсах и возможностях, предоставляемых сервисом, основанной на спецификациях OGC-совместимых интерфейсов.

В рамках интеграции распределенных пространственных данных и геоинформационных систем различных источников одной из важнейших задач является обеспечение эффективного информационного обмена между такими системами, а также эффективного поиска и доступа к данным, где ключевую роль играют метаданные.

При этом ввиду интеграции большого количества независимых геоинформационных систем существенной проблемой является различная трактовка структуры и семантики информации в разных системах. Данные могут относиться к различным предметным областям, в рамках одной иметь разные выражение и интерпретацию. Наиболее эффективным способом обеспечения согласованного информационного обмена в таких условиях является обеспечение так называемой семантической интероперабельности систем на основе стека технологий семантического веба: RDF, RDFS, OWL. Язык Web-онтологий OWL представляет развитые возможности для описания предметных областей взаимодействующих систем, определения используемых терминов и их взаимосвязи. В данной работе рассматриваются принципы применения технологий Semantic Web и OWL-онтологий для обеспечения

семантической интероперабельности обмена пространственными ресурсами в распределенной среде Интернета, приводятся принципы формирования наборов онтологий-стандартов для описания метаданных пространственных данных и сервисов.

Методология создания профиля пространственных метаданных с помощью онтологического подхода обеспечивает преодоление трудностей, присущих классическим подходам к профилям метаданных, выраженным в UML и схемах XML, которые недостаточно гибки, чтобы принять, отклонить или переопределить элементы метаданных. Преимущество методологии состоит в том, что она не требует дублирования спецификаций. Каждый элемент в спецификации метаданных и каждый термин предметной области однозначно определен с использованием URI. Онтологический подход кодирования метаданных использует объектно-ориентированные особенности, такие как наследование, для осуществления связи профилей семейств с расширяемой спецификацией метаданных.

В результате анализа различных современных решений и технологий, связанных с публикацией и поиском пространственных ресурсов в Интернете, проведенного автором работы, и на основе общих технических требований, предъявляемых к реализации геопортала и системы управления пространственными метаданными как ключевой компоненты инфраструктуры пространственных данных, сформулированы технические и функциональные требования к геоинформационно-аналитическим системам на основе Web-портала. На основе анализа проблем взаимодействия компонентов распределенных ГИС-систем обоснована необходимость разработки специализированной онтологической модели пространственных данных и сервисов на основе единой (базовой) схемы метаданных.

Глава II. Стандарты, кодировки и онтологический подход к проблеме моделирования пространственных метаданных

Во второй главе приводится анализ спецификаций пространственных метаданных большинства существующих на данный момент ИПД и геоинформационных порталов, которые были вовлечены в исследовании для формирования академического профиля пространственных метаданных. Особое внимание уделяется в анализе таким стандартам и профилям пространственных метаданных, как международный стандарт ISO 19115:2003 «Географическая информация. Метаданные», американский стандарт FGDC-STD-001-1998, российский стандарт ГОСТ Р 52573–2006 «Географическая информация. Метаданные», стандарт CEN (prEN 12657 Geographic Information – Metadata), который получил наибольшее распространение в Европе и австралийский и новозеландский профиль ИПД ANZLIC, профиль ИПД Великобритании – UK GEMINI и другие.

Далее в главе освещены сильные и слабые стороны различных типов кодирования спецификаций метаданных. Очевидно, что в настоящее время стандарт ISO 19115 является лучшей разработкой в области пространственных метаданных, и именно на него следует ориентироваться при построении ИПД и различных геоинформационных систем. Подобные стандарты присутствуют во всех национальных ИПД, в частности и в России в виде ГОСТа Р 52573–2006 «Географическая информация. Метаданные». К сожалению, отличия национальной версии ГОСТ Р 52573–2006 от международной весьма значительны, так что метаданные, подготовленные в соответствии с национальным профилем, не будут совместимы с международным стандартом. Российским разработчикам придется реализовывать два варианта средств подготовки метаданных — российский и международный, а также создавать и использовать «национальные» конверторы для зарубежного программного обеспечения ГИС, применяемого в России. Это идет вразрез с основной идеей ИПД — упростить обмен пространственными данными. В связи с этим, разработка или адаптация спецификации метаданных, которая помогла бы описать формат и содержание данных для достижения максимальной interoperability, определяют актуальность исследования данной работы.

Кроме ISO 19115:2003, важную роль играет недавно утвержденный международный стандарт ISO 19139:2007 «Geographic information -- Metadata -- XML schema implementation», который предлагает схему XML-кодирования для описания, подтверждения и обмена метаданными в ISO 19115:2003. Если каталог пространственных данных поддерживает кодировки данных XML, кодированных в соответствии с ISO 19139, то он будет обладать полной функциональной совместимостью и сможет поддерживать обмен данными с любым построенным на стандартах ISO каталогом.

В данной главе описывается подход к интеграции данных геоинформационных систем, основанный на использовании онтологии пространственных данных и сервисов, задающей определения понятий предметной области и связей между ними. Один из практических подходов к представлению онтологий заключается в их определении в виде словаря понятий предметной области, отношений между ними и накладываемых на них ограничений.

Представленный подход обеспечивает процессу формирования академического профиля метаданных максимальную полноту и совместимость для реализации формальных онтологий пространственных данных и сервисов, используемых на последующих шагах конструирования разрабатываемой информационной системы.

Глава III. Формирование академического профиля пространственных метаданных и разработка онтологии пространственных данных и сервисов

Третья глава посвящена формированию академического профиля метаданных «ГеоМЕТА» и реализации онтологии пространственных данных и сервисов. Несмотря на рекомендации ISO/TC211, FGDC и OGC на использование стандартов пространственных метаданных, проведенный в главе 2 анализ показывает, что достижение согласия по созданию одной единственной спецификации пространственных метаданных в мировом масштабе практически невозможно.

В рамках проведенного исследования анализируются международные и российские спецификации пространственных метаданных и предлагается простой, но в то же время достаточно полный профиль метаданных для пространственных данных и сервисов, ориентированный на максимальную совместимость, адаптируемость и расширяемость как для российских, так и для зарубежных поставщиков и потребителей геоинформационных ресурсов.

Особое внимание в процессе формирования профиля «ГеоМЕТА» уделяется решению проблемы совместимости стандарта ГОСТ Р 52573–2006 с международным стандартом ISO 19115 и предложен подход к решению этой проблемы, позволяющий сохранить без изменений элементы и коды, описанные в международном стандарте, дополнив их национальными. В случае конфликтов кодов словарных терминов необходимо поддерживать оба набора кодов. В итоге зарубежные средства подготовки метаданных могли бы использоваться для создания и ведения базовых элементов метаданных (в большинстве случаев этого достаточно), а программные продукты российского производства были бы пригодны для подготовки метаданных в соответствии с международным стандартом.

В результате анализа для разработки схемы метаданных были взяты стандарт ISO 19115 и скорректированный российский стандарт ГОСТ Р 52573–2006 для описания пространственных данных. На основе наследования элемента MD_Identification из стандарта ISO 19119 введен элемент SV_ServiceIdentification, который агрегирует ещё три класса SV_OperationMetadata, SV_Parameter, SV_ServiceProvider, служащих для описания сервисов, ориентированных на спецификации сервисов OGC и веб-сервисы.

На практике 22 элементов ядра ISO 19115 оказывается недостаточно для описания ресурсов. Для максимальной совместимости к ядру ISO 19115 были добавлены ещё 14 элементов для данных и 12 элементов для описания сервисов. В итоге ядро «ГеоМЕТА» v3.0:2007 для пространственных данных и сервисов состоит из 48 элементов, а всего в профиле имеется 226 элемента (см. рисунок 1).

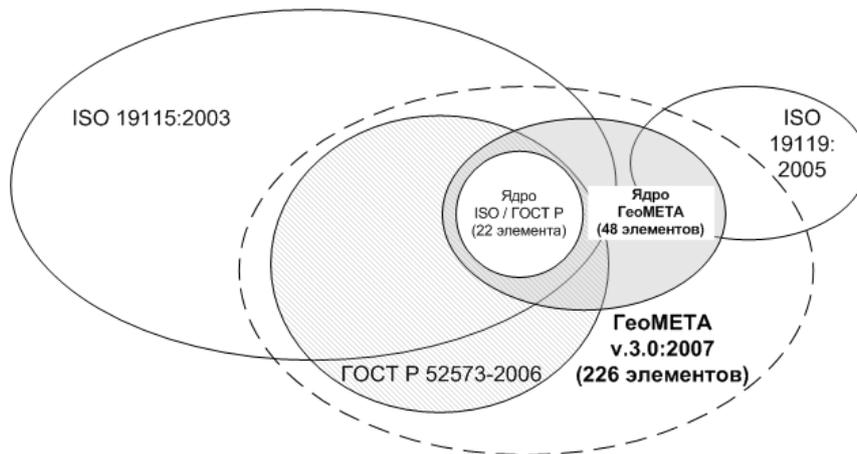


Рис. 1. Академический профиль пространственных метаданных.

Для моделирования схемы и структуры используется язык UML. Полное описание разработанного профиля приводится в приложении. Основные пакеты метаданных и графическое отображение отношений между отдельными пакетами представлены на рисунке 2.

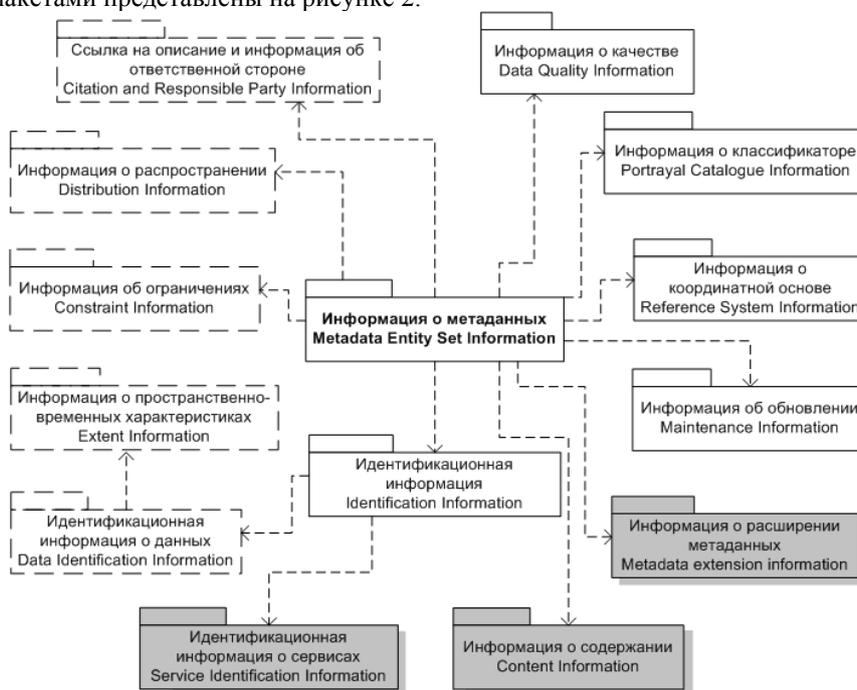


Рис. 2. Пакеты академического профиля пространственных метаданных

Всего в профиле содержится 14 пакетов, из которых 3 являются новыми (на рисунке 2 отмечены серым цветом). Остальные были наследованы от ГОСТ Р 52573-2006 с учетом модификации в 5-и пакетах (на рисунке 2 отмечены пунктирной линией) до уровня совместимости с ISO 19115:2003 и расширения важных элементов для описания сервисов и улучшения поиска.

Задача интеграции распределенных неоднородных источников геоинформационных данных решается путем выделения единой (базовой) схемы метаданных, реализованной в виде OWL-онтологии на основе разработанного академического профиля метаданных «ГеоМЕТА» для обеспечения «взаимопонимания» (т.е. семантической интероперабельности) систем, участвующих в распределенном взаимодействии.

Разработанная онтология пространственных данных и сервисов и подробное описание профиля метаданных «ГеоМЕТА» доступны по адресу <http://earth.ras.ru>. Онтологическая модель была описана на языке OWL DL, количество строк составило 5221, а размер файла – 366 Кбайтов, в нее вошло 48 классов, количество понятий составило 209. Пример части онтологической модели пространственных данных и сервисов приведен на рисунке 3. Ниже кратко описываются четыре основных класса, используемые в онтологии.

Для каждого набора ресурса, который регистрируется системой, создается соответствующий экземпляр класса модели данных или сервисов, который сохраняется в базе знаний.

Идентификация данных (MD_DataIdentification): Этот класс позволяет поставщику специфицировать идентифицирующую информацию набора данных. Важными атрибутами этого класса являются: title (название), abstract (описание), topicCategory (тематика), descriptiveKeywords (совокупность ключевых слов), creator (создатель), pointOfContact (субъект, ответственный за создание), resourceMaintenance (частота осмотра), resourceConstraints (ограничение) и др. Многие атрибуты этого класса являются субатрибутами мета-элементов Dublin Core(DC) (в версии OWL), которые предоставляют стандарт для описания информационных ресурсов.

Идентификация сервисов (SV_ServiceIdentification): Данный класс позволяет поставщику специфицировать идентифицирующую информацию набора сервисов. Важными атрибутами этого класса являются: serviceType (тип сервиса, например: WMS, WFS, WCS и веб-сервисы), operationName (название операции), parameterName (название параметров), connectionPoint (точка доступа) и др. Этот класс использует онтологии OGC Service Architecture, NASA GCMD Service Type, разработанные исследовательской группой GeoBrain.

Содержимое данных (MD_ContentInformation): Данный класс является ключевым в онтологии модели данных и отвечает за отображение набора данных в концепции предметной области, определенные в геонаучной онтологии. Эта связка генерирует семантическую концептуальную схему для набора данных. Поставщик данных и сервисов выбирает уровни детализации из геонаучной онтологии, которая наилучшим образом описывает набор данных. Этот набор хранится в классе содержимого данных, позволяя, таким образом, онтологии модели данных предоставлять не только метаданные о наборе данных, но и семантическое описание содержимого набора данных.

Результатом интеграции станет единая среда, позволяющая осуществлять распределенный поиск пространственных данных и сервисов, визуализировать их, получать к ним доступ и обрабатывать необходимым образом. Эта онтологическая модель определяет элементы для включения информации об идентификации данных и сервисов, временных и пространственных характеристиках, ограничениях, визуальном представлении, собственно содержательной части и распространении данных в зависимости от набора данных и т.д. Эти элементы отвечают на вопросы «кто, что, почему, где, когда и как» каждого аспекта набора ресурсов. При комбинировании онтологий, относящихся к другим областям, с предлагаемой онтологией появляются новые возможности для создания семантических метаданных для наборов данных, что демонстрируется на примере выполненной реализации геоинформационно-аналитической системы «ГеоМЕТА» с пользовательским Web-интерфейсом геопортала. Система позволяет поставщику пространственных

данных и сервисов выбирать уровни детализации из онтологий, которые будут наилучшим образом описывать метаданные ресурсов. Эти метаданные могут быть использованы конечными пользователями пространственных ресурсов для улучшенного поиска по содержанию и оценки пригодности независимо от того, где ресурсы находятся и в каком формате они хранятся.

Глава IV. Проектирование, программная реализация и апробация геоинформационно-аналитического Web-портала на основе пространственных онтологий пространственных данных и сервисов

Четвертая глава описывает на основе разработанной онтологии пространственных данных и сервисов программное средство геоинформационно-аналитической системы «ГеоМЕТА», представляющее собой платформу для интеграции распределенных неоднородных источников геоинформационных данных и сервисов, доступных в мире. Особый акцент делается на обеспечении их семантической интероперабельности вне зависимости от способа их представления. Система предоставляет единую точку входа (Web-портал) к этой среде, которая позволит пользователем в сфере наук о Земле:

- удобно и гибко интегрировать/распространять распределенные пространственные данные и сервисы на уровне (мета)данных в соответствии с основными международными и российскими стандартами;
- высокий уровень интероперабельности, масштабируемости и гибкости;
- легко находить специализированные данные и приложения (OGC сервисы и веб-сервисы);
- доступ к ресурсам через Интернет по стандартизованным протоколам OGC и W3C;
- производить вычислительные эксперименты;
- визуализировать результаты деятельности.

Система «ГеоМЕТА» состоит из двух частей, в совокупности обеспечивающих решение поставленных задач – части по работе с пространственными (мета)данными и геосервисам (ГИС-части, непосредственно являющейся предметом данной работы) и части по работе с приложениями для обработки пространственных данных (научно-вычислительного портала). Описание и реализация части научно-вычислительного портала (НВП) не являются предметом исследования данной диссертационной работы, а представляют собой предмет другого исследования, выполняющегося в рамках данной разработки.

К функциональностям ГИС-части системы относятся:

- каталогизация пространственных метаданных;

- локальный и распределенный поиск пространственных данных и сервисов на основе метаданных;
- автоматизированный сбор распределенных пространственных метаданных на основе OGC WMS/WFS сервисов;
- интеграция пространственных данных и (гео)сервисов участников системы (всего 3 типа участников);
- предоставление доступа к распределенным пространственным данным по стандартизированным протоколам;
- визуализация карт, редактирование элементов.

Функциональность Научно-вычислительного портала в рамках системы «ГеоМЕТА» сужается на область ГИС-приложений и обеспечивает:

- каталогизацию «вычислительных приложений», атрибутивный поиск вычислительных приложений (ВП) в каталоге;
- запуск ВП, мониторинг выполнения и визуализация результатов;

конструирование и сохранение рабочих процессов из существующих ВП.

Интерфейс системы представлен Web-порталом, поэтому для ГИС-части основным методом доступа пользователя к информации является обычный доступ к Web-страницам портала через любой распространенный браузер (см. рисунок 4) по адресу <http://earth.ras.ru>. Ядро системы представляет платформу ЕНИП и обеспечивает следующие возможности:

- управление статическим содержанием;
- хранение объектов системы (представленных RDF-тройками) в РСУБД;
- индексирование и полнотекстовый поиск;
- обеспечение безопасности системы.

В рамках данной работы не рассматривается подробное описание функциональности ЕНИП и её возможности, отметим лишь используемые технологии непосредственно для реализации системы «ГеоМЕТА».

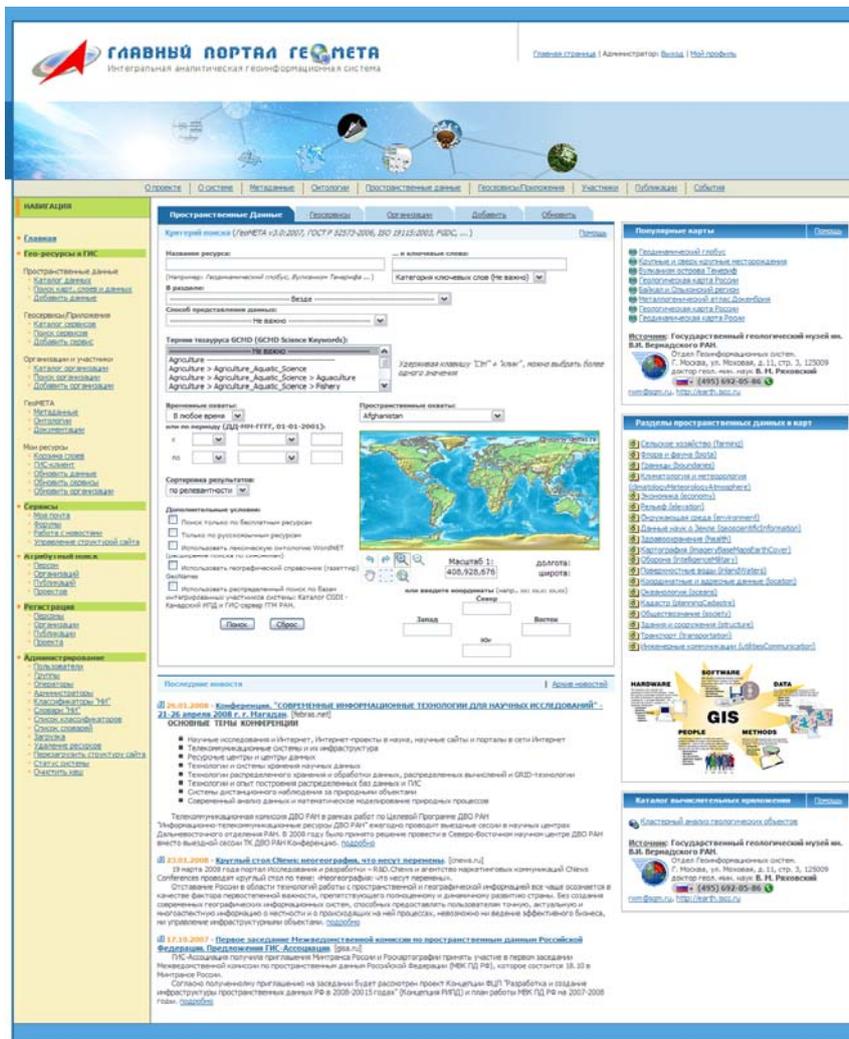


Рис. 4. Главная страница веб-портал «ГеоМЕТА».

Архитектура системы «ГеоМЕТА» приведена на рисунке 5 и включает основные сервисные компоненты и сервер данных «ГеоМЕТА», каталог ресурсов и онтологии, ядро системы – платформа ЕНИП, а также схемы интеграции ресурсов 3-х типов участников и взаимодействия пользователей с

системой. Системные блоки, которые являются предметом исследования и реализации данной работы, отдельно отмечены серым фоном.

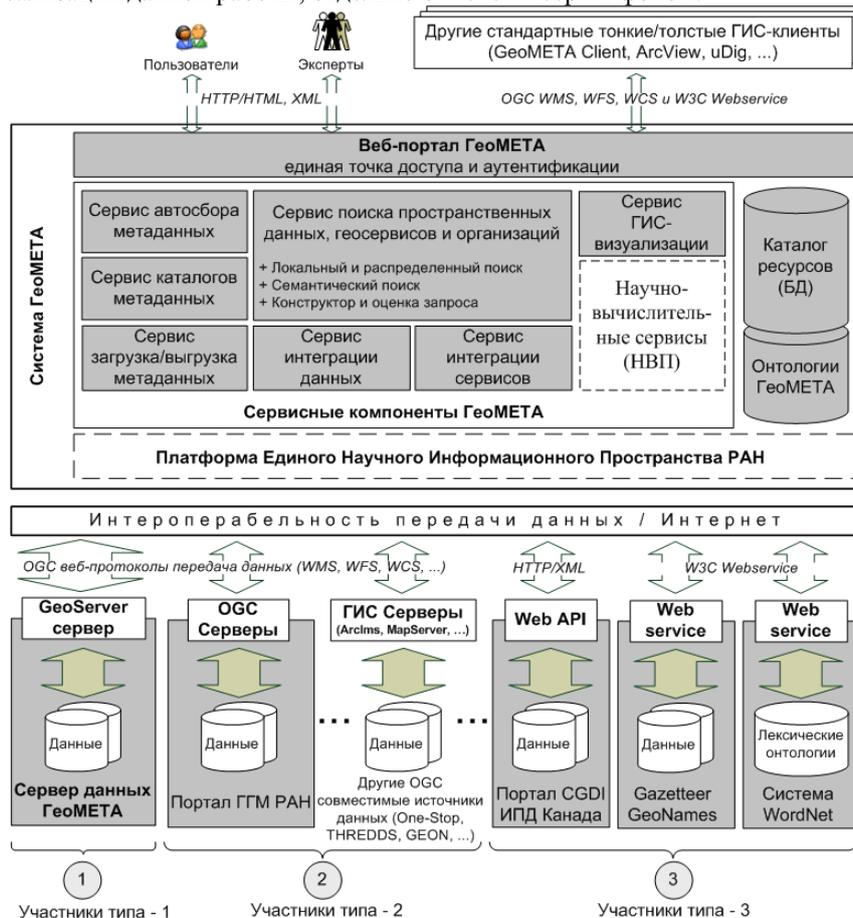


Рис. 5. Архитектура системы «GeoMETA».

Архитектура системы разрабатывалась на основе тщательного анализа большинства существующих на данный момент ИПД и геоинформационных порталов. Основой программной реализации системы «GeoMETA» стала система «Научный институт РАН», а созданная онтология пространственных данных расширила схему метаданных ЕНИП РАН, что обеспечило возможность включения портала системы «GeoMETA» в ЕНИП РАН. При реализации ГИС-части системы использовались открытые решения, такие как MapBuilder, GeoServer и т.п.

Заключение

В заключении приведены основные результаты диссертационной работы.

Приложение

В приложении приведены диаграммы классов, описывающие основные сущности онтологической модели пространственных данных и сервисов, их атрибуты и связи между ними, а также примеры формирования пространственных метаданных на основе разработанных онтологий. В приложении приведено также полное описание 226 элементов разработанного академического профиля пространственных метаданных.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

1. Предложены подходы семантического описания пространственных метаданных и онтологии геоданных и сервисов для решения семантической проблемы разделения, поиска и обмена пространственными данными и сервисами.
2. Разработан академический профиль и OWL-онтология пространственных метаданных с поддержкой двуязычного словаря терминов с описаниями для интеграции распределенных наборов пространственных данных и сервисов, совместимых с международным стандартом ISO 19115:2003, ISO 19119:2005 и ГОСТ Р 52573-2006 в совокупности с набором программных и пользовательских веб-интерфейсов для регистрации/редактирования метаданных и доступа к его данным на основе протокола HTTP по стандартам OGC Web Map Service (WMS), Web Feature Service (WFS), включая возможность ГИС-визуализации.
3. Предложены и реализованы схемы поиска и многоуровневая каталогизация распределенных пространственных данных и сервисов и автоматизированного сбора пространственных метаданных на основе OGC-совместимых сервисов.
4. Реализован программный комплекс, включающий геоинформационный Web-портал, для проведения научных исследований на основе распределенных пространственных данных и сервисов – одна из первых в России онлайн-систем управления пространственными метаданными, нацеленная на поддержку интероперабельности между стандартами ГОСТ Р 52573-2006 и ISO 19115:2003 (а также совместимая со схемами метаданных, лежащих в основе наиболее известных в мире ИПД) с возможностью обмена метаданными в структурированном XML-формате по стандарту ISO 19139:2007.
5. Разработанные онтологии, методы и технологии успешно использованы в ряде геоинформационных проектов отдела систем математического обеспечения ВЦ РАН.

СПИСОК ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

1. *Динь Ле Дат, Серебряков В.А.* Разработка и реализация формальных онтологий геопространственных данных и сервисов // Журнал "Радиотехника". – 2008. – вып.2 – С. 85-89.
2. *Динь Ле Дат.* Интеграция GRID и GIS технологий // Наука & Сотрудничество между Вьетнамом и Россией. VIII международная научная конференция: Сборник трудов научной конференции /Вьетнамская научно-техническая ассоциация в РФ. – М., 2006. – С. 177-187.
3. *Динь Ле Дат.* Общий обзор ГИС и стандарты пространственных метаданных // Наука & Сотрудничество между Вьетнамом и Россией. IX международная научная конференция: Сборник трудов научной конференции /Вьетнамская научно-техническая ассоциация в РФ. – М., 2007. – С. 157-165.
4. *Dinh Le Dat, Vershinin A.V., Serebriakov V.A.* Scientific Computational Web Portal - Based Distributed GIS – Enabled GRID // The 12th International GIS-net Conference and Exhibition on GIScience, Ho Chi Minh, Viet Nam. – 2006. – P. 14-15.

В работах с соавторами личный вклад автора заключается в создании академического профиля и OWL-онтологии метаданных пространственных данных и сервисов, поддерживающие одновременное полное совместимо с российским стандартом ГОСТ Р 52573–2006 и совместимо с международными стандартами ISO 19115:2003, ISO 19119:2005. Автором предложен онтологический подход к описанию модели пространственных ресурсов с использованием технологий Semantic Web, описана архитектура Интернет-системы интеграции распределенных неоднородных источников геоинформационных ресурсов, являющейся ГИС-частью главного геопортала системы «ГеоМЕТА» ВЦ РАН, создан соответствующий комплекс программ для каталогизации и поиска пространственных ресурсов и автоматизированного сбора метаданных на основе OGC-совместимых сервисов.