

MultiMeta - Система интеграции пространственных данных и ресурсов электронных библиотек *

© О.М. Атаева, А.А. Каленкова, В.А. Серебряков

Учреждение Российской академии наук
Вычислительный центр им. А.А.Дородницына РАН, г.Москва

oli@ultimeta.ru, akalenkova@ultimeta.ru, serebr@ccas.ru

Аннотация

В статье представлена Система интеграции пространственных данных и ресурсов электронных библиотек MultiMeta. Приводится схема метаданных MultiMeta, построенная на базе схемы ЕНИП [16], она включает основной и библиотечный профили ЕПИП, а также схемы для работы с пространственными данными, музейными и медиа-объектами, которые были разработаны в рамках систем GeoMeta [15] и LibMeta [19]. Дается описание основных функциональных и интеграционных возможностей MultiMeta.

1. Введение

Пространственные данные – это цифровые данные о пространственных объектах, включающие сведения об их местоположении и свойствах, пространственных и непространственных атрибутах [21].

В последние годы идет поиск механизмов, которые позволили бы различным организациям работать вместе в направлении разделения и обмена своими пространственными данными и ресурсами для предоставления эффективных информационных сервисов различным пользователям [5]. Это соображение обуславливает внесение ряда инициатив. Одна из них: Инфраструктура Пространственных Данных (ИПД) (Spatial Data Infrastructure, SDI) берет свое начало с ранних девяностых [4], когда предложение по развитию национальной инфраструктуры пространственных данных получило общественное признание, признание частного сектора и сообщества пользователей. В соответствии с определением, представленным в [17, 4], ИПД – информационно-телекоммуникационная система, обеспечивающая

доступ граждан, хозяйствующих субъектов, органов государственной и муниципальной власти к распределенным ресурсам пространственных данных, а также распространение и обмен данными в общедоступной глобальной информационной сети в целях повышения эффективности их производства и использования. С развитием Интернета, Всемирной паутины и географических информационных систем (ГИС) [20, 22] приложения, в которых используются распределенные системы географической информации, получили существенное развитие. Распределенная географическая информация представляет собой всеобщее распространение географической информации во множестве форм, включая карты, снимки, наборы данных, операции по анализу и отчеты [10]. Интеграция средств Интернета и географических информационных систем открывает новые области исследований.

Кроме того, в силу специфики своей предметной области участники процесса интеграции обладают данными и метаданными, дополняющими пространственную информацию: это могут быть публикации, включающие набор метаданных и полный текст, музейные экспонаты, представленные метаданными и медиа-объектами, самостоятельные медиа-объекты, такие как фотографии, видео и аудиозаписи. Интеграция разнородной информации, предоставляемой несколькими источниками, позволит максимально составить описание территории, провести анализ, выявить взаимосвязи, получить принципиально новые возможности поиска: ресурсы, даже хранящиеся в разных системах, представляются связанными друг с другом единой системой навигации.

Системой, которая позволяет интегрировать пространственные и другие (в частности библиотечные и музейные) метаданные, используя стандартные форматы и протоколы обмена, и является представленная в этой работе Система интеграции пространственных данных и ресурсов электронных библиотек – MultiMeta.

Труды 13^й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011, Воронеж, Россия, 2011.

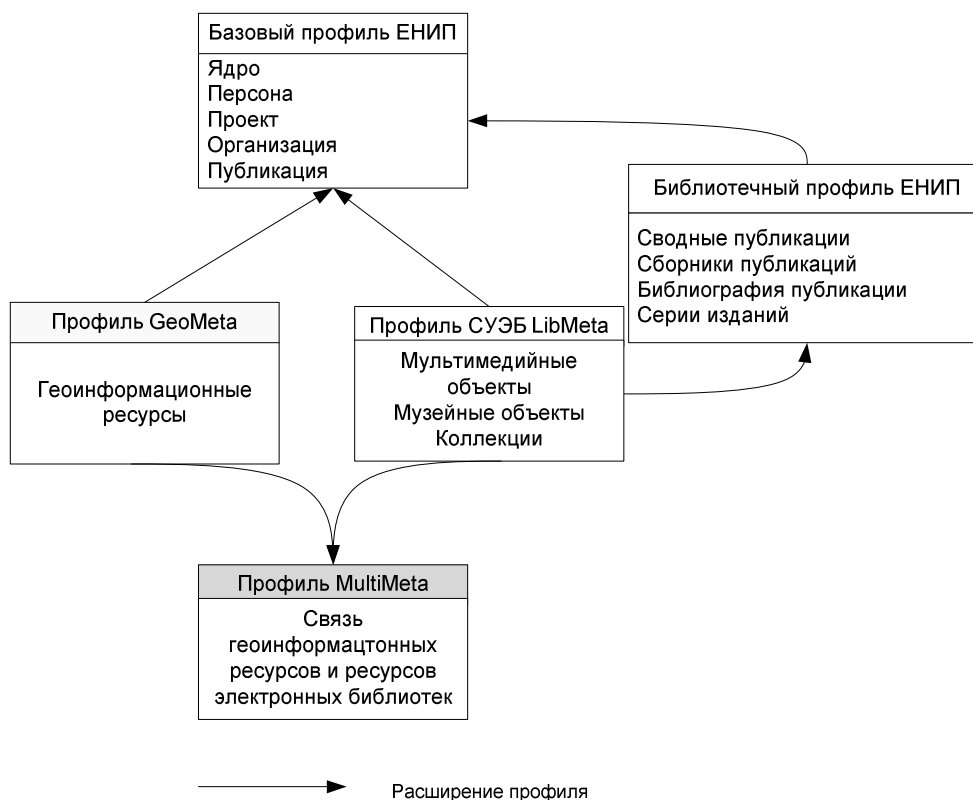


Рис. 1. Схема метаданных MultiMeta

2. Схема метаданных MultiMeta

Схема метаданных Системы MultiMeta построена на базе схем Портала GeoMeta и Системы управления электронными библиотеками LibMeta (СУЭБ LibMeta), которые в свою очередь проектировались на основе архитектуры Единого Научного Информационного Пространства (ЕНИП). Схема метаданных, которая составляет основу MultiMeta, представлена на рисунке 1. MultiMeta наследует такие преимущества ЕНИП, как ориентация на семантический Веб (Semantic Web) [1], возможность уникальной идентификации информации, возможность поиска информации в распределённой среде, возможности интероперабельности с внешней средой.

Построенная на основе портала GeoMeta, Система MultiMeta естественным образом наследует схему ЕНИП и возможность работы с такими ресурсами как «персона», «публикация», «проект» и «организация». Кроме того, центральным элементом схемы для MultiMeta остается «геоинформационный ресурс». Первоначально «геоинформационный ресурс» обладает атрибутами, представляющими идентификационную информацию, информацию о качестве, координатной основе, распространении, карте и слоях. Схема информационных ресурсов, служащих дополнением «геоинформационному ресурсу», таких как тексты, видео и аудиозаписи, полностью представлена в СУЭБ LibMeta, поэтому она включена в схему MultiMeta. Для возможности

работы с библиографическими описаниями и сборниками публикаций в MultiMeta включен Библиотечный профиль ЕНИП, непосредственно из профиля метаданных СУЭБ LibMeta наследуются схемы для работы с мультимедийными и музейными объектами, коллекциями публикаций. Благодаря тому, что сами схемы метаданных формализованы с помощью стандартов RDF [11]/RDFS [12] / OWL [9], они легко объединяются и дополняются: сам профиль метаданных MultiMeta включает набор объектных свойств, которые связывают «геоинформационные ресурсы» и такие ресурсы как «коллекция», «музейный объект» (связь «геоинформационного ресурса» и «публикации» уже реализована в GeoMeta, связь «мультимедийного объекта» и произвольного «ресурса» представлена в LibMeta). Таким образом, интеграция пространственных данных и ресурсов электронных библиотек базируется на унификации и совместимости схем метаданных.

3. Функциональные возможности MultiMeta

Система интеграции пространственных данных и ресурсов электронных библиотек MultiMeta включает в себя следующие функциональные подсистемы, доступные пользователю через web-интерфейс:

- Подсистема работы с метаданными об ученых, публикациях, музейных объектах и

геоинформационных объектах позволяет просматривать, редактировать, а также производить поиск информации об ученом, публикации, музейном объекте.

- Подсистема работы с коллекциями позволяет просматривать, редактировать и выполнять поиск по коллекции.
- Подсистема работы с наборами дополнительных атрибутов дает возможность создавать наборы атрибутов, назначать их некоторому музейному предмету.
- Подсистема работы с медиа-объектами позволяет просматривать и редактировать медиа-объекты.
- Подсистема хранения и просмотра отсканированных текстов дает следующие возможности: просматривать подряд страницы издания, переходить на любую заданную страницу (в том числе на предыдущую, на последующую, на страницу с заданным номером), просматривать оглавления издания (с возможностью перехода на нужный раздел), просматривать страницы в увеличенном масштабе, выполнять разворот иллюстраций на 90°.
- Подсистема управления структурой статического наполнения портала.
- Подсистема управления группами и пользователями.
- Подсистема управления новостями.
- Подсистема ведения словарей и классификаторов, которые могут быть использованы для организации тематического поиска.
- Подсистема пакетной загрузки данных позволяет загружать данные в формате RDF/XML в соответствии с онтологической моделью метаданных MultiMeta.
- Подсистема полнотекстового поиска информации об ученых, публикациях, музейных объектах, коллекциях и медиа-объектах.
- Подсистема каталогизации, сбора, поиска пространственных метаданных.
- Подсистема предоставления доступа к пространственным данным по стандартизованному протоколу.
- Подсистема визуализация карт.
- Подсистема получения метаданных по протоколу OAI-PMH [8], в частности получение и загрузка метаданных публикаций и геоинформационных ресурсов в формате Dublin Core [13].

Кроме того, были созданы страницы просмотра списков медиа-объектов, коллекций, музейных объектов, связанных с геоинформационным

ресурсом, а также страницы привязки указанных типов объектов к геоинформационному ресурсу.

4. Интеграционные возможности MultiMeta

Система MultiMeta позволяет интегрировать как пространственные данные с помощью подсистем, реализованных в Системе GeoMeta и «унаследованных» MultiMeta, так и ресурсы электронных библиотек.

Задача интеграции распределенных неоднородных источников пространственных данных решается путем выделения базовой схемы метаданных, реализованной в виде OWL-онтологии на основе стандартов ISO 19115:2003 «Географическая информация. Метаданные» [6] и ISO 19139:2007 «Географическая информация. Метаданные – XML схема» [7] и полностью включающей российский стандарт ГОСТ Р 52573-2006 «Географическая информация. Метаданные» [18] (а также учитывающей профили стандартов и схемы метаданных, лежащие в основе наиболее известных национальных инфраструктур пространственных данных), для обеспечения семантической интероперабельности систем, участвующих в распределенном взаимодействии. На данный момент возможна загрузка метаданных в формате ISO 19139:2007.

Предусмотрена возможность периодического сбора метаданных (harvesting) из различных источников – каталогов метаданных. При этом система 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 из удаленного источника и их синхронизацию с заданной периодичностью.

В настоящее время реализованы два протокола обмена: GeoNetwork – порталы на базе ресурсов системы «Geonetwork opensource» [3] и Web DAV [14] – каталоги на базе сервера WebDAV, используются для обмена с другими порталами, кроме GeoNetwork. В обоих случаях текст метаданных передается в формате XML по стандарту ISO 19139:2007.

Для получения пространственных и непространственных метаданных в MultiMeta полностью реализованы стандарты OAI-PMH и Dublin Core. Кроме того, для интеграции с музейными системами существует отображение метаданных системы на концептуальную модель CIDOC-CRM [2]. Для использования других протоколов и форматов обмена данными и метаданными, в особенности основанных на XML,

не представляется сложной реализация модулей обмена.

Таким образом, в MultiMeta наряду с загрузкой геоинформационных ресурсов может быть выполнена загрузка публикаций, музейных и медиа-объектов.

Связь загружаемых объектов и пространственных данных может быть указана при загрузке в формате RDF/XML, соответствующем схеме метаданных MultiMeta. Рассматривается вопрос об указании связи при загрузке в формате Dublin Core.

5. Заключение

В работе представлена созданная авторами Система MultiMeta, позволяющая интегрировать пространственные данные и ресурсы электронных библиотек, основываясь на известных стандартных форматах и протоколах. Кроме того, Система MultiMeta построена на основе единой схемы метаданных и позволяет загружать информацию в формате RDF/XML в соответствии с онтологической моделью. В дальнейшие планы авторов входит улучшение дизайна системы и ее наполнение данными мультимедийного электронного атласа Курильских островов и других электронных атласов, предоставляемых Институтом географии РАН.

Литература

- [1] Berners-Lee T., Shadbolt N., Hall W. The Semantic Web Revisited // IEEE Intelligent Systems. 2006. N. 6.
- [2] CIDOC CRM Home page. <http://cidoc.ics.forth.gr/index.html>.
- [3] Geonetwork opensource. The complete manual. V 2.4, 2009. <http://geonetwork-opensource.org/>.
- [4] Groot R., McLaughlin J. Geospatial Data Infrastructure Concepts, Cases and Good Practices / NY: Oxford University Press, 2000.
- [5] Guarin J. M. M., Radwan M. M. Architecting distributed geo-information services beyond data infrastructures // XXth ISPRS congress: Geoimagery bridging continents.
- [6] ISO 19115:2003 – Geographic information. Metadata.
- [7] ISO 19139:2007 – Geographic information – Metadata – XML schema implementation.
- [8] Open archives initiative protocol for metadata harvesting. <http://www.openarchives.org/pmh>.
- [9] OWL Web Ontology Language Semantics and Abstract Syntax. <http://www.w3.org/TR/2004/REC-owl-semantics-20040210/>.
- [10] Plewe B. GIS Online Information Retrieval, Mapping, and the Internet / USA: Onward Press, 1997.
- [11] Resource Description Framework (RDF) Model and Syntax. <http://www.w3.org/TR/rdf-primer/>.
- [12] Resource Description Framework (RDF) Schema Specification. <http://www.w3.org/TR/rdf-schema>.
- [13] The Dublin Core Metadata Element Set: an American national standard. <http://dublincore.org/documents/dces/>.
- [14] WebDAV Resources. <http://www.webdav.org/>
- [15] Атаева О.М., Кузнецов К.А., В.А. Серебряков, Филиппов В.И. Портал интеграции пространственных данных «ГеоМета» // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. 2010. – С. 219.
- [16] Бездушный А. Н., Бездушный А. А., Серебряков В. А., Филиппов В. И. Интеграция метаданных Единого Научного Информационного Пространства РАН / М.: ВЦ РАН, 2006.
- [17] ГОСТ Р 52438-2005 – Географические информационные системы. Термины и определения.
- [18] ГОСТ Р 52573–2006 – Географическая информация. Метаданные. <http://protect.gost.ru/document.aspx?control=7&id=74833>.
- [19] Захаров А.А., Серебряков В.А. Система управления электронными библиотеками LibMeta // Электронные библиотеки: перспективные методы и технологии, электронные коллекции. 2010. – С. 28.
- [20] Майкл Н. ДеМерс. Географические информационные системы основы / М: Дата+, 1999.
- [21] Словарь по естественным наукам. Глоссарий.ru http://slovari.yandex.ru/dict/gl_natural/article/97/097_214.HTM.
- [22] Томлинсон, Роджер Ф. Думая о ГИС. Планирование географических информационных систем: руководство для менеджеров. Пер. с англ. / М.: Дата+, 2004.

MultiMeta – a System of Spatial and Digital Library Resources Integration

© О.М. Атаева, А.А. Kalenkova, V.A. Serebriakov

The system of spatial and digital library resources integration – MultiMeta is presented. MultiMeta metadata schema is given; it is based on ENIP schema and includes the main and library profiles of the ENIP. Also it contains schemas to deal with geospatial data, museum and media objects which were developed within GeoMeta and LibMeta systems. Main functional and integration capabilities of MultiMeta are exposed.

* Работа выполняется в рамках проекта РФФИ №11-07-00286-а.