# СИСТЕМА УПРАВЛЕНИЯ ЭЛЕКТРОННОЙ БИБЛИОТЕКОЙ LIBMETA*

*Серебряков В.А., Филиппов В.И., Каленкова А.А.*
*(Вычислительный центр им. А.А. Дородницына РАН)*

С 2007 года в ВЦ РАН ведутся работы по созданию СУЭБ в рамках ЕНИП под названием LibMeta, которая предлагает библиотекам, архивам и музеям РАН унифицированное решение, позволяющее публиковать полные тексты научных работ и разнообразные мультимедийные материалы, интегрируясь существующими информационными системами РАН при соответствии стандартам в области ЭБ. Портал ЭБ «Научное наследие России» является первой установкой СУЭБ LibMeta и площадкой для обкатки технологических и архитектурных решений. В докладе на примере портала ЭБ «Научное наследие России» представлены общая архитектура, профиль метаданных и интеграционные возможности СУЭБ LibMeta.

## Введение

В последние годы объемы информации в сети Интернет в связи с бурным ее развитием лавинообразно увеличиваются [1]. Несмотря на все большее проникновение технологий Semantic Web [2, 3], ощущается серьезная нехватка средств поиска и каталогизации информации, которые позволяли бы искать ее именно по семантике и связям, а не только по ключевым словам и полным текстам, как это делают универсальные поисковые системы. Одним из способов решения данной проблемы видится появление и все большее распространение различного рода электронных библиотек (ЭБ) [4, 5].

Интеграция ЭБ с любыми (и не только библиотечными) ресурсами обеспечивает отсутствие дублирования данных: данные могут храниться в одной центральной информационной системе, при этом в других системах находятся ссылки на эти данные. Если исходные данные

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

располагаются в различных системах, то они могут быть реплицированы в центральную информационную систему и автоматически обновляться при обновлении оригинала. Кроме того, возможно хранение в центральной информационной системе лишь метаинформации, необходимой для навигации и семантического поиска, в то время как сами данные будут располагаться в других информационных системах. При этом ресурсы, даже хранящиеся в разных системах, представляются связанными друг с другом единой системой навигации. Такой единой информационной системой, реализующей указанные подходы к интеграции данных, и является система управления электронными библиотеками LibMeta [6].

## Профиль метаданных СУЭБ LibMeta

Профиль метаданных СУЭБ LibMeta построен на основе профиля метаданных Единого научного информационного пространства (ЕНИП) [7]. В профиле метаданных ЕНИП для электронных библиотек используются ресурсы, такие как Организации, Персоны, Публикации.

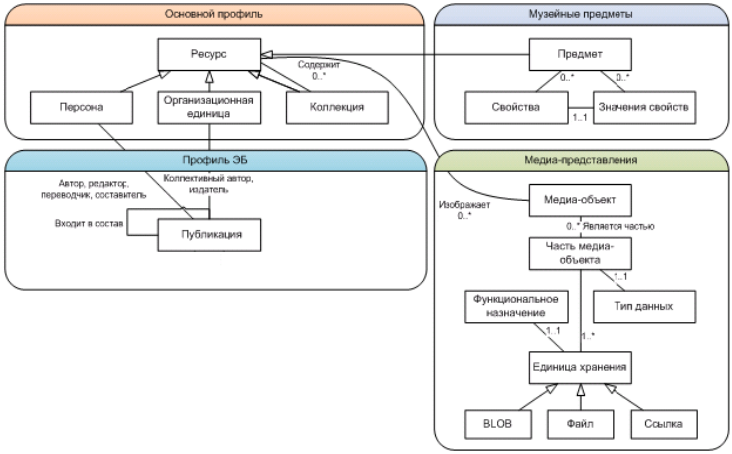


*Рис. 1. Схема профилей метаданных СУЭБ LibMeta*

В целях обеспечения поддержки различных уровней детализации информации о публикациях, необходимых различным приложениям, библиографическая специализация разделена на базовую и расширенную подсхемы, а также выделяется академическая подсхема, отражающая специфику научных публикаций. Уже на базовом уровне требуется структурировать информацию обо всех вышестоящих библиографических уровнях для каждой публикации. Например, для описания ряда статей в журнале, необходимо описать сам журнал как издание сводного уровня, далее описать интересующие выпуски этого журнала как издания монографического уровня, и, наконец, сами статьи как издания аналитического уровня. И статья, и выпуск, и журнал как таковой являются полноценными структурированными ресурсами, описываемыми лишь единожды, и связываемыми с помощью URI (Unified Resource Identifier) — ссылок.

Такой структурированный подход требует некоторого усилия со стороны систем с «планарным» описанием публикаций. Однако, структуризация информации обо всех библиографических уровнях необходима и крайне важна для схем электронных библиотек. Она позволяет избежать дублирования информации, эффектов наличия опечаток в названиях группирующих выпусков, серий и пр., позволяет представить пользователю информацию в целостном и непротиворечивом виде. Общая схема профилей метаданных, применяемых в СУЭБ LibMeta, а также основных сущностей в данных профилях приведена на рисунке 1.

К основным типам данных, представленных в СУЭБ LibMeta, относятся Публикации, Персоны (авторы), Предметы. Сближение задач электронных библиотек, архивов и музеев выдвигает требование стандартизации метаданных физических музейных предметов и их мультимедийных (фото, видео, аудио) представлений. В связи с этим в СУЭБ LibMeta разработаны дополнительные прикладные профили поддержки музейной деятельности и мультимедийных представлений.

В отличие от публикаций, описания музейных объектов могут значительно отличаться в различных музеях и здесь невозможно обеспечить всеобъемлющий набор не-

обходимых свойств. В связи с этим для данных объектов реализуется возможность определения дополнительных свойств в виде связей с двумя вспомогательными объектами: Дополнительные свойства и Значения дополнительных свойств. Соответственно, в интерфейсе администратора системы предоставляется возможность определять дополнительные свойства предмета, при этом в интерфейсах ввода и вывода данных создаются представления соответствующих полей. Введенные значения дополнительных полей выдаются в полных сведениях о предмете, но поиск по ним не производится. Таким образом, администратор может добавить такие свойства, как Количество предметов, Автор описания, География, Размеры, Возраст, Способ поступления, Препараты и т.п.

Для обеспечения цифровых представлений не только публикаций, но и музейных объектов, а также мультимедийных изображений коллекций, фотографий и т.п., вводится ряд новых сущностей, в класс Ресурс, являющийся суперклассом для всех основных объектов онтологии, вводится свойство Медиа-представление. Таким образом, одно или несколько мультимедийных представлений могут сопровождать любой объект информационной системы, наследуемый от класса Ресурс.

В основном профиле метаданных ЕНИП предусмотрена поддержка коллекций, однако требования цифровых библиотек, а в особенности с поддержкой хранения музейных предметов, не позволяют их полноценно использовать. В связи с эти базовый профиль дополняется коллекциями со следующими атрибутами: Название, Тип коллекции (элемент словаря), Ключевые слова, Описание, Администратор (ссылка на Персону), Количество элементов в коллекции, Место хранения, Примечание, Элементы коллекции (ссылка на Ресурс). Коллекции такого рода позволяют хранить классические коллекции (архивные, музейные) и иметь любые вложенные наборы объектов (выставочные, выездные и пр.).

## Общая архитектура СУЭБ LibMeta

Система управления электронной библиотекой LibMeta включает в себя следующие функциональные подсистемы:

- Подсистема работы с метаданными об ученых, публикациях, музейных 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 возможность 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; обеспечивает возможность просмотра страниц в увеличенном масштабе, выполнять разворот иллюстраций на 90º.
- 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 [8] в соответствии с онтологической моделью метаданных LibMeta.
- Подсистема полнотекстового поиска информации об ученых, публикациях, музейных объектах, коллекциях и медиа-объектах.

- Подсистема импорта метаданных, а также подготовленных электронных изданий и их оглавлений из внешних систем.

В настоящее время на Портале ЭБ «Научное наследие России», являющимся установкой СУЭБ LibMeta, импорт метаданных персон и публикаций с Сервера подготовки метаданных (БЕН РАН) выполняется по протоколу HTTP. При этом, метаданные не проверяются на наличие дубликатов в системе СУЭБ LibMeta, так как Сервер подготовки метаданных (БЕН РАН) пока является единственным поставщиком информации о публикациях и персонах. Тем не менее, подсистема импорта метаданных из произвольных внешних информационных систем также поддерживает получение метаданных по протоколу OAI-PMH [9] и проверку на наличие дубликатов. Опишем работу этой системы подробнее.

**Интеграция СУЭБ LibMeta с другими информационными системами**

В системе создан универсальный модуль загрузки метаданных в некотором XML-формате в соответствии с протоколом OAI-PMH. Алгоритм получения метаданных некоторого ресурса, реализованный в этом модуле, представлен на рисунке 2. С определенной периодичностью интеграционный модуль запрашивает вновь созданные или измененные метаданные из удаленного хранилища по протоколу OAI-PMH. В первую очередь, проверяется URI получаемых метаданных. Если метаданные с указанным URI уже представлены в системе, то выполняется XSLT [10] — преобразование (метаданные приводятся к внутреннему RDF/XML формату СУЭБ LibMeta) и производится загрузка в режиме «дозапись». При загрузке в режиме «Дозапись новых данных поверх существующих», для каждого свойства, загружаемого из RDF/XML, все прежние значения этого свойства стираются и заменяются на значения из RDF/XML. При этом значения тех свойств, которые были указаны в базе, но отсутствуют в RDF/XML, оставляются неизменными. Такой режим загрузки обеспечивает корректную инкрементную «дозапись» данных поверх существующих. Если метаданных с указанным URI в системе нет, то они являются

новыми, и также должны быть загружены. Однако, в силу того, что СУЭБ LibMeta представляет собой единый интеграционный узел, метаданные, соответствующие некоторому информационному ресурсу, могли быть получены ранее из другого источника. Для того чтобы в СУЭБ LibMeta не возникало дубликатов, используется вспомогательный модуль автоматической проверки на дубликаты [11]. Если есть предположение о том, что загружаемые метаданные уже хранятся в системе, источнику метаданных отправляется информация о схожих метаданных, находящихся в СУЭБ LibMeta.

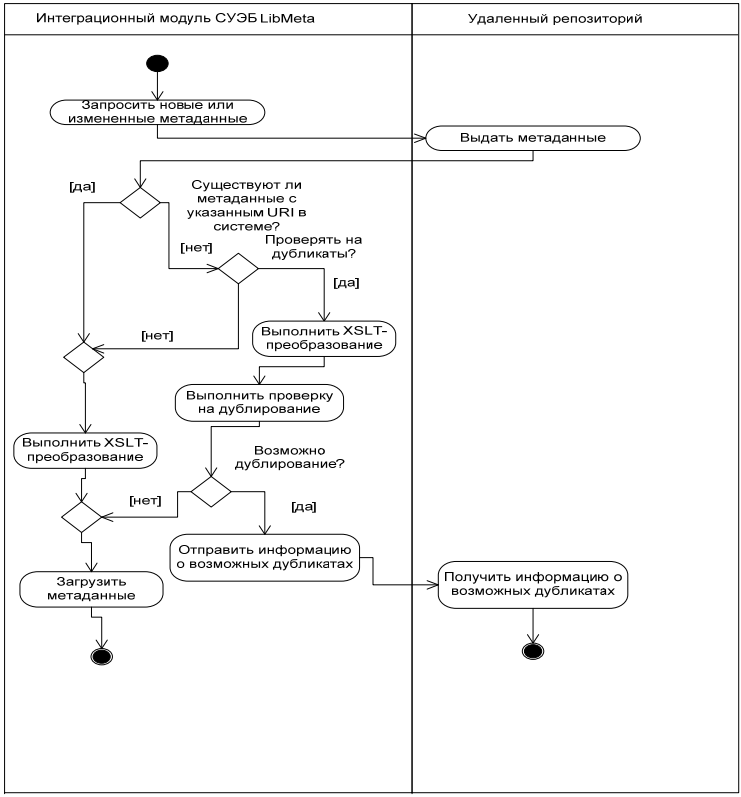


*Рис. 2. Алгоритм работы интеграционного модуля СУЭБ LibMeta.*

На стороне источника определяется, соответствуют ли метаданные одному и тому же информационному ресурсу. Если принимается решение о том, что эти метаданные уже есть в системе, для них устанавливается URI уже загруженных метаданных (тогда при следующей загрузке метаданные в репозитории могут быть дополнены новыми значениями полей), иначе для них выставляется признак того, что они должны быть загружены, несмотря на наличие схожих метаданных, и они попадают в систему при следующей загрузке без проверки на дублирование. Таким образом, интеграционный модуль СУЭБ LibMeta реализует некоторый общий подход к загрузке метаданных из удаленного репозитория.

*Литература*


1. *Gantz J., Chute C., Manfrediz A., et al. Доклад IDC при финансовой поддержке компанией EMC: Обновленный прогноз роста мирового объема информации до 2011 г.*
2. *Berners-Lee T., Hendler J., Lassila O. The Semantic Web // Scientific Am., 2001. N. 5. P. 34–43.*
3. *Berners-Lee T., Shadbolt N., Hall W. The Semantic Web Revisited // IEEE Intelligent Systems, 2006. N. 6.*
4. *Галева И. С. Интернет как инструмент библиографического поиска. — М.: Профессия, 2007.*
5. *Зацман И. М. Концептуальный поиск и качество информации. — М.: Наука, 2003.*
6. *Захаров А.А., Серебряков В.А. Система управления электронными библиотеками LibMeta // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XII Всероссийской научной конференции RCDL'2010. — Казань: КФУ, 2010. — 28 с.*
7. *Филиппов В.И. Захаров А.А. Поддержка цифровых библиотек и музейных объектов в среде ЕНИП // Информационное обеспечение науки. Новые технологии Сб. науч. тр. / Каленов Н.Е. (ред.). — М.: Научный Мир, 2009. — с. 113-124.*

8. *Resource Description Framework (RDF) Model and Syntax, W3C Recommendation, 2004. — http://www.w3.org/ TR/rdf-primer/.*
9. *Open archives initiative protocol for metadata harvesting. — http://www.openarchives.org/ pmh.*
10. *XSL Transformations (XSLT) Version 2.0, W3C Recommendation, 2007. — http://www.w3.org/TR/xslt20/.*
11. *Атаева О.М., Шиолашвили Л.Н. Методы очистки интегрируемых данных // Современные проблемы фундаментальных и прикладных наук: Труды XLIX научной конференции. / Моск. физ.-тех. ин-т. — М., 2006.*