

Система управления электронными библиотеками LibMeta

© А.А. Захаров, В.А. Серебряков

ВЦ РАН, г. Москва

andrey@sufler.ru, serebr@ccas.ru

Аннотация

Представлена система управления электронной библиотекой (СУЭБ) LibMeta, предназначенная для научных институтов РАН, входящих в ЕНИП. Описана её архитектура, рассмотрены стандарты, положенные в её основу, а также решаемые ею основные проблемы. Также предлагается вниманию электронная библиотека «Научное наследие России», основанная на СУЭБ LibMeta.

1 Введение

В течение последних лет, в связи с бурным развитием интернета и экспоненциальным увеличением количества информации в нем, всё больше и больше ощущается потребность в наличии средств поиска и каталогизации информации, которые позволили бы находить информацию не только по ключевым словам, как это делают универсальные поисковые системы, но и по семантике и отношениям её с другой информацией. Одной из мер по удовлетворению такой потребности является появление и всё большее распространение различного рода электронных библиотек (ЭБ).

Зачастую понятие ЭБ смешивают с уже давно существующими электронными каталогами – информационными системами, предназначенными для использования в обычных библиотеках, содержащими только метainформацию и служащими средствами поиска. ЭБ – это совершенно другой класс систем: они хранят не только метаданные, но и полные тексты информационных ресурсов. Ещё одним важным отличием ЭБ от электронных каталогов является то, что в качестве информационного ресурса может пониматься фактически всё, что угодно, например, музейные предметы или архивные материалы. В связи с этим можно говорить о сближении задач ЭБ, электронных архивов и цифровых музеев – все эти информационные системы фактически являются ЭБ с различной специализацией.

Однако на пути развития ЭБ как в мире, так и в России до сих пор есть множество препятствий. Главным из них является отсутствие стандартизации в данной области. Конечно же, существует большое количество стандартов, так или иначе

касающихся разных аспектов ЭБ, однако нет такого всеобъемлющего стандарта, который описывал бы такие системы со всех точек зрения; кроме того, в настоящее время во многих системах ЭБ зачастую стандарты не применяются вовсе либо применены в ограниченном или изменённом виде.

В Вычислительном центре РАН на протяжении нескольких лет ведётся исследование вопросов, связанных с ЭБ, и как результат этих исследований в данной статье представлена СУЭБ LibMeta.

В первой части статьи более подробно рассмотрены основные проблемы и требования к ЭБ, во второй части обсуждаются наиболее полезные стандарты в сфере ЭБ и в третьей части представляется СУЭБ LibMeta.

2 Требования к электронной библиотеке

Современный мир предъявляет повышенные требования ко всем информационным системам: с точки зрения пользователей они должны быть удобными в использовании, простыми в изучении; а с технической точки зрения должны быть тесно связаны с другими информационными системами и предоставлять стандартизированные службы. Все эти требования в полной мере относятся и к электронным библиотекам.

Мы будем рассматривать требования к ЭБ в контексте ЕНИП, поскольку СУЭБ LibMeta разрабатывалась как СУЭБ, в первую очередь предназначенная для научных институтов в составе РАН и других организаций, желающих предоставить свои ресурсы в ЕНИП.

2.1 Интеграция с информационными системами

Фактически любая современная информационная система и, в частности, ЭБ должна быть интегрирована с другими информационными системами. Такой подход даёт сразу несколько серьёзных преимуществ:

- отсутствие дублирования данных: исходные данные хранятся только в одной информационной системе, в других системах они используются по ссылкам того или иного вида либо реплицируются и автоматически обновляются при обновлении оригинала;
- централизованные сервисные службы, например, служба аутентификации и авторизации

пользователей: пользователи системы вводят свою идентификационную информацию и проходят аутентификацию единожды для всей группы связанных информационных систем;

- интеграция информационных ресурсов: ресурсы, даже хранящиеся в разных системах, представляются связанными друг с другом единой системой навигации.

Естественно, чтобы снизить сложность разработки таких интегрированных информационных систем, имеет смысл строить их на одной общей архитектурной и технологической основе. Одним из таких архитектурных решений является единое научное информационное пространство (ЕНИП) РАН. Оно призвано удовлетворить потребность научных сотрудников в необходимости как поиска качественной информации, так и выставления собственной информации в интернете.

Основу ЕНИП РАН составляют, прежде всего, стандарты на метаданные информации, циркулирующей в ЕНИП [1]. Эти стандарты должны отвечать следующим требованиям:

- включать в себя основные типы информации, требующейся для поддержки работы научно-го сотрудника;
- быть открытыми, то есть обеспечивать доступ к соответствующей информации по этим описаниям;
- быть расширяемыми, то есть обеспечивать возможность детализации описаний;
- обеспечивать возможности интеграции информации;
- обеспечивать возможности уникальной идентификации информации;
- обеспечивать возможности размещения и поиска информации в распределённой среде;
- обеспечивать возможности интероперабельности с внешней средой.

Кроме перечисленного, рекомендуется, чтобы схемы метаданных были ориентированными на семантический веб (Semantic Web) [2, 3].

Схемы метаданных играют в ЕНИП двоякую роль. С одной стороны, они служат «обменными схемами», с разными уровнями детализации, для обмена данными между системами, входящими в ЕНИП. С другой стороны, в рамках ЕНИП стоит задача не только предложить обменные схемы, но и разработать конкретные типовые информационные системы для научных институтов, библиотек, издательских отделов и пр., которые дали бы стимул к информационному наполнению ЕНИП. Каждую конкретную предметную область предлагается описывать отдельной схемой, точнее, набором схем (профилем метаданных), соответствующим образом ссылающихся друг на друга.

2.2 Распределённость

Одним из наиболее бурно развивающихся направлений информационной индустрии последних

лет стала разработка распределённых информационных систем. Причинами такого бурного роста стали достижения одновременно в нескольких областях, среди них отметим следующие:

- значительный рост пропускной способности каналов связи; скорость обмена по ним приближается к скоростям внутренних шин компьютеров;
- рост производительности компьютеров как по скорости, так и по объёмам памяти, и оперативной, и внешней;
- широкое проникновение компьютеров и компьютерных технологий в повседневную деятельность как большинства организаций и учреждений, так и граждан;
- развитие интернета, обеспечивающего простой и надёжный доступ к невероятному числу информационных ресурсов;
- развитие самих информационных технологий; с этой точки зрения можно сказать, что программирование находится на четвёртой фазе своего развития: 1) «классическое» программирование (сначала в кодах, затем на ассемблере, затем на языках высокого уровня) для больших ЭВМ; 2) «классическое» программирование для персональных ЭВМ; 3) программирование с использованием визуальных и CASE-средств; 4) «сетевое» программирование.

Тенденция по увеличению доли распределённых систем не обошла и Российскую академию наук, которая имеет разветвлённую структуру, объединяет большое число научно-исследовательских учреждений и коллективов, расположенных на всей территории России и вовлечённых во все многообразие видов научной деятельности. Эти учреждения обладают уникальными научными информационными ресурсами, среди которых: опубликованные результаты научных исследований и экспериментов, библиографические и фактографические базы данных, сведения об учёных, их научной деятельности, публикациях, проектах и т. п. Эти ресурсы представляют значительный интерес для сотрудников научных и административных учреждений, членов мирового научного сообщества, представителей промышленности и предпринимателей, которые заинтересованы во внедрении результатов научных исследований.

В связи с перечисленными факторами при разработке современной ЭБ, в особенности, в среде ЕНИП, повышенное внимание следует уделять созданию средств взаимодействия ЭБ друг с другом, а также средств организации распределённых ЭБ, в которых данные, объём которых зачастую может быть настолько велик, что их нельзя хранить в одной ЭБ, «размазаны» между отдельными хранилищами. При этом существуют специальные средства поиска и каталогизации, работающие над всеми данными такой распределённой ЭБ.

2.3 Следование стандартам

Когда говорят о стандартизации в сфере ЭБ, зачастую приходят к трём основным проблемам:

- отсутствие единых взглядов на архитектуру ЭБ;
- наличие большого количества стандартов на форматы метаданных и данных;
- слабая реализация стандартов в существующих ЭБ.

Оправданием обычно служит то, что каждая ЭБ по своей сути уникальна, и все они не могут опираться на одни и те же стандарты. Тем не менее, следует отметить, что в мире обычных библиотек метаданные (картотека) достаточно хорошо формализованы и стандартизованы, по крайней мере на уровне национальных стандартов. Таким образом, можно считать, что необходим по крайней мере единый и достаточный стандарт на метаданные.

Со стандартизацией форматов данных ситуация гораздо сложнее: в настоящее время существует огромное количество «стандартов» их представления. В связи с этим возникает вопрос, в каком формате данные должны храниться в ЭБ и должны ли они в том же самом формате предоставляться конечным пользователям. Зачастую оказывается, что формат, удобный для хранения в хранилищах данных (обеспечивающий целостность данных, лёгкость доступа к ним, содержащий минимальный набор метаданных), не слишком удобен для пользователей, поскольку требует от них установки специального программного обеспечения, либо не приспособлен к передаче по сети, так как не обеспечивает необходимой степени сжатия. Таким образом, возникает необходимость поддержки как минимум двух форматов представления данных для каждого вида ресурсов (аудио, видео, статические изображения). Также не следует забывать, что с течением времени существующие стандарты устаревают и возникают новые, что приводит к необходимости перехода на новые форматы.

Что касается единого архитектурного стандарта, то здесь существует два основных предложения по архитектуре ЭБ: DELOS DLRM и OAIS, рассмотренные далее. Эти стандарты покрывают разные области архитектуры ЭБ и могут быть использованы совместно, однако следует отметить, что они не описывают ни логическую, ни тем более физическую модель ЭБ, поэтому результат их применения в разных реализациях ЭБ может сильно различаться.

2.4 Подготовка ресурсов

В большинстве случаев, когда ЭБ не предназначена для размещения так называемых «исходно цифровых» (born-digital) ресурсов, подготовка ресурсов к публикации является достаточно сложным и трудоёмким процессом, вовлекающим

многих участников. За простотой пользовательского интерфейса скрывается целая подсистема подготовки ресурсов, включающая в себя следующие компоненты:

- службы оцифровки – службы, создаваемые при библиотеках, музеях и других поставщиках данных, там, где возможен непосредственный доступ к оцифровываемым материалам;
- служба контроля качества оцифровки – единая служба контроля качества оцифровываемых данных, в которой задействованы специалисты по оцифровываемым предметам; несмотря на то, что первичный контроль качества выполняется при самой оцифровке, зачастую происходит утеря какой-либо части оцифровываемых данных либо, с точки зрения специалиста по оцифровываемым данным оцифровка выполнена в не полном объёме;
- служба подготовки метаданных – иногда может быть объединена со службой оцифровки, однако следует учесть, что ввод метаданных должен осуществлять не специалист по оцифровке, а специалист в предметной области;
- служба окончательной подготовки ресурсов и контроля качества, которая выполняет объединение данных и метаданных, следит за обновлением данных и метаданных и осуществляет общий контроль качества подготавливаемых ресурсов.

Перечисленные службы должны обладать своими информационными системами, обеспечивающими выполнение их задач и предоставляющими друг другу данные, необходимые для работы. В результате работы подсистемы подготовки ресурсов получают готовые к публикации в ЭБ ресурсы, которые по каналам связи автоматически представляются к публикации.

3 Стандартизация в электронных библиотеках

Проблемой стандартов для электронных библиотек занимаются многие сообщества и организации, но, несмотря на значительные усилия и имеющиеся достижения, единства мнений добиться пока не удаётся. Необходимость стандартизации в этой области ощущается, пожалуй, сильнее, чем для других видов информационно-поисковых систем, так как многие проекты электронных библиотек стараются объединить ресурсы нескольких существующих библиотечных и/или архивных систем.

Стандарты, имеющие отношение к ЭБ, условно можно разделить на три большие группы:

- архитектурные стандарты, описывающие принципы и крупные компоненты, применяемые при построении ЭБ;
- стандарты метаданных и их представления, описывающие наборы метаданных и, воз-

можно, способы их представления при передаче и хранении;

- стандарты информационного обмена, описывающие протоколы передачи метаданных и данных между системами в различных целях.

3.1 Архитектура электронных библиотек

Архитектурные стандарты описывают, как должна быть устроена ЭБ. Такие описания могут касаться как интерфейсов взаимодействия с другими системами и пользователями, так и состава и назначения компонентов системы. Также стандарты этой группы пытаются дать чёткое определение самому термину ЭБ и всех связанных с ним терминов. Также к данной группе могут относиться и стандарты и предложения, связанные с ЭБ, но напрямую не посвящённые им.

3.1.1 DELOS DLRM

Группа DELOS [4] (существует под различными названиями с 1996 года) является одной из наиболее известных и старых организаций, занимающихся стандартизацией в области ЭБ. Основными направлениями деятельности организации являются исследование и создание стандартов в области архитектуры ЭБ, их интеграции и отчуждения данных.

Самым интересным результатом работы DELOS является образцовая модель ЭБ – Digital Library Reference Model (DELOS DLRM) [5]. В данной модели большое внимание уделяется отделению цифровых объектов (содержимого ЭБ) от ЭБ и от СУЭБ (рис. 1). Также вводятся различные классы пользователей ЭБ, решающие свои задачи в рамках ЭБ, которым ЭБ (и СУЭБ) должна предоставлять соответствующую функциональность, приводится весьма обширная концептуальная модель данной области, с тщательными определениями важнейших представлений об архитектуре, ресурсах и функциональности ЭБ, в частности, в UML-представлении.



Рис. 1. СУЭБ, ЭБ и содержимое

3.1.2 OAIS

Open Archival Information System (OAIS) [6] – это разрабатываемый и поддерживаемый организацией Consultative Committee for Space Data Systems (CCSDS) [7] международный стандарт (ISO 14721:2003) на образцовую модель информа-

ционной системы открытого архива. В нём описывается подход к долговременному хранению данных, в частности, освещается проблема устаревания форматов данных и физических носителей информации. Также описываются схемы взаимодействия участников архивной системы при различных сценариях работы. Кроме технической стороны работы открытого архива, рассматривается также работа административно-управленческих отделов при ЭБ, занимающихся планированием хранения информации, предоставлением доступа и другими задачами.

3.2 Представление метаданных

Стандарты метаданных играют очень важную роль – они описывают, в каком формате данные передаются из одной информационной системы в другую. Следует заметить, что не все стандарты метаданных кроме самого описания сущностей, их связей и атрибутов представляют описание контейнера для метаданных, однако большая часть стандартов подразумевает использование XML как одного из видов контейнеров. В данном разделе описаны как универсальные стандарты метаданных, такие, как Dublin Core, так и специализированные, например, CIDOC-CRM.

3.2.1 Dublin Core

Стандарт Dublin Core [8, 9] состоит из двух частей – «минимальной» Dublin Core Metadata Element Set (DCES), имеющей статус международного стандарта (ISO 15836:2009), и полной – DCMI Metadata Terms. Обе версии стандарта не привязаны ни к какой конкретной предметной области и могут описывать ресурсы любых видов. Минимальная версия содержит только 15 атрибутов, применяемых к любым сущностям, в которых можно описать название объекта, его описание, автора, административные метаданные и ссылку на источник. Полная версия содержит все атрибуты минимальной, а также набор дополнительных сущностей для описания различных атрибутов, набор словарей и дополнительное множество атрибутов. Чаще всего используется именно минимальный набор, поскольку именно он позволяет максимально абстрагироваться от предметной области, но при этом, конечно же, проигрывая в деталях.

Также существует рекомендация DCMI Abstract Model, содержащая сведения по расширению DCMI Metadata Terms для нужд различных предметных областей.

Кроме описания схемы метаданных стандарты Dublin Core содержат описание контейнеров, в частности, описываются формат DC-Text для хранения в чисто текстовом формате, формат DC-XML для XML-представления и DC-RDF для представления в RDF-подмножестве XML.

3.2.2 CIDOC-CRM

С 1994 года в составе International Council of Museums (ICOM) существует комитет Committee on Documentation of the International Council of Museums (ICOM-CIDOC) [10], который занимается стандартизацией в области музейных метаданных. Основной целью данного комитета является создание образцовой концептуальной модели (CRM) для описания сущностей и связей, используемых в документировании культурного наследия. В 1999 году была выпущена первая версия CIDOC-CRM [11], а в 2006 модель получила статус международного стандарта (ISO 21127:2006).

Модель CIDOC-CRM чрезвычайно подробно описывает предметную область и поэтому на практике в полном объёме практически никогда не применяется. Тем не менее, отображение метаданных ЭБ на эту модель представляется полезным, так как многие музейные системы заявляют свою совместимость с CIDOC-CRM.

3.2.3 FRBR и FRBRoo

Изначально разрабатывавшийся The International Federation of Library Associations and Institutions (IFLA) [12] стандарт Functional Requirements for Bibliographic Records (FRBR) [13] играет примерно такую же роль в области описания публикаций, какую имеет CIDOC-CRM в описаниях предметов культурного наследия. С 2000 года при содействии комитета CIDOC разрабатывается объектно-ориентированный стандарт FRBRoo [14], заимствующий часть сущностей и идей из CIDOC-CRM.

3.2.4 PRISM

Стандарт Publishing Requirements for Industry Standard Metadata (PRISM) [15] разработан издательскими организациями, входящими в IDEAlliance (International Digital Enterprise Alliance) для обмена метаданными о публикациях. PRISM основан на DCMI, но в большей степени ориентирован на библиографические ресурсы. В стандарте предлагаются среда обмена и хранения данных и метаданных и ряд словарей значений этих элементов. В ЕНИП используются схема контролируемых словарей PRISM и основной набор элементов PRISM.

3.2.5 MARC

Стандарт MARC (MAchine-Readable Cataloging) по праву может считаться одним из самых старых стандартов в информационных технологиях. Он был разработан в начале 1960 годов в Библиотеке Конгресса США. Основное предназначение стандарта – хранение библиографических записей для электронных каталогов. В настоящее время оригинальный MARC (USMARC) уже не используется, на смену ему пришли моди-

фикации MARC 21 [16] (США и Канада) и национальные модификации (например, RUSMARC [17] в России). Также существует отображение MARC на XML, называемое MARCXML [18]. Хотя MARC основан на устаревших технологиях, он достаточно широко используется, и многие современные информационные системы заявляют свою с ним совместимость.

3.3 Стандарты информационного обмена

В последнее время в связи с общей направленностью на децентрализацию и распределённость всё большую значимость приобретают стандарты информационного обмена. К ним относятся стандарты распределённого поиска, репликации метаданных и репликации данных.

3.3.1 OAI-PMH

Стандарт Open Archives Initiative – Protocol for Metadata Harvesting [19] является признанным лидером среди стандартов распределённого поиска и репликации метаданных. В основу такой популярности положено то, что он достаточно прост для реализации и может инкапсулировать метаданные в любом XML-формате. Каждая OAI-PMH-совместимая система должна поддерживать метаданные в формате Dublin Core (см. п. 3.2.1).

3.3.2 OAI-ORE

Object Reuse and Exchange [20] – Semantic-Web ориентированный стандарт описания и обмена агрегированными ресурсами. Стандарт описывает способы объединения частей объекта (данных, метаданных) в один агрегированный ресурс, доступный по единому URL, а также способы обработки и предоставления частей объекта по такому URL.

4 СУЭБ LibMeta

С 2007 года в ВЦ РАН ведутся работы по созданию системы управления ЭБ [5] под названием «СУЭБ LibMeta», которая позволила бы библиотекам, архивам и музеям РАН иметь унифицированное решение, позволяющее размещать полные тексты публикаций и разнообразные мультимедийные материалы, быть интегрированной с существующими информационными системами РАН, а также соответствовать стандартам в области ЭБ.

Портал проекта «Научное наследие России» [21] является первой установкой СУЭБ LibMeta, а также площадкой для обкатки технологических и архитектурных решений.

4.1 Архитектура СУЭБ LibMeta

Поскольку предполагается использование СУЭБ LibMeta в научных институтах различной направленности, невозможно предоставить одну

единственную схему метаданных, подходящую абсолютно под все задачи. Данная проблема может быть решена двумя способами: внесением избыточности в схему метаданных и предоставлением администратору системы возможности доопределять схему. Оба подхода не лишены своих недостатков. Так, например, избыточная схема приводит к тому, что в каждой конкретной установке ЭБ используется только часть схемы, зачастую достаточно небольшая. Это приводит к нерациональному использованию ресурсов СУБД и, в целом, к уменьшению быстродействия. Кроме того, внесение такой избыточности требует глубоких исследований в предметной области, а также наличия общепринятых стандартов на подобные метаданные. Второй подход также не лишён недостатков: основным является то, что по метаданным, определённым администратором, поиск может осуществляться только с большими затратами ресурсов. Кроме того, уменьшаются возможности по связыванию ресурсов друг с другом, так как администратор может определять только содержательные атрибуты.

В СУЭБ LibMeta применяются оба подхода: для ресурсов типа публикация и персона применяется избыточная схема, разработанная на основе библиотечных стандартов, а для музейных предметов используется второй подход. Такое решение основывается на том, что для музейных предметов практически невозможно создать единую (избыточную) схему, поскольку каждый музей обладает своей спецификой. Естественно, что для музейных предметов существует и фиксированная часть схемы, позволяющая связать их с другими ресурсами и включающая основные содержательные метаданные.

При разработке СУЭБ LibMeta были проанализированы мировые стандарты в области ЭБ а также обобщён опыт их использования в других библиотечных системах. На основе этого анализа были выдвинуты предложения по использованию стандартов. В качестве архитектурных стандартов были использованы как DELOS [5], так и OAIS [6]. Из DELOS взяты общие концепции и варианты использования ЭБ. Из OAIS почерпнуты сведения о процессах, происходящих внутри ЭБ, а также её взаимодействии с внешними системами. Из стандартов метаданных были использованы Dublin Core (unqualified) [9], CIDOC-CRM [11], RUSMARC [17] и другие. На эти стандарты существует отображение схемы метаданных СУЭБ LibMeta и возможен информационный обмен по протоколам, поддерживающим инкапсуляцию данных в этих форматах. Перечисленные стандарты являются наиболее распространёнными в мире ЭБ и их использование позволяет достичь максимальной совместимости с существующими системами.

В качестве основного стандарта информационного обмена выбран OAI-PMH [19], поскольку

он поддерживается большей частью библиотечных и архивных систем.

Основой для обмена данными и семантической интероперабельности в ЕНИП служат технологии Semantic Web, соответственно, существенным является применение в ЕНИП существующих предложений по стандартизации наборов элементов метаданных для Semantic Web. В технологиях Semantic Web широко используются язык RDF, а также его специализация для описания онтологий OWL. Логично было выбрать именно OWL как язык описания метаданных в СУЭБ LibMeta, а RDF – как язык обмена метаданными между системами.

В качестве инфраструктурного решения для реализации СУЭБ LibMeta была выбрана программная платформа «Научный Институт» [22], разработанная ВЦ РАН. Эта платформа представляет собой типовой программный комплекс автоматизации информационной деятельности научного института в составе Российской академии наук, обеспечения научной деятельности его сотрудников, взаимодействующий с другими информационными системами в составе ЕНИП. «Научный Институт» изначально разрабатывается как модульная расширяемая система, позволяющая гибко подбирать наиболее подходящий набор функциональных возможностей для каждой конкретной организации. Основу информационной системы составляют ядро системы («платформа»), программное решение, предназначенное для создания распределённых информационных систем, веб-порталов и интеграции данных, а также набор функциональных модулей, предоставляющих специальные функции. Для СУЭБ LibMeta «Научный Институт» является слоем абстракции от СУБД и контейнера приложений. Кроме того, в СУЭБ LibMeta используется библиографический профиль метаданных ЕНИП, предоставляющий большую часть схемы метаданных в части определения таких ресурсов, как публикации, персоны, проекты и организации. Также в состав «Научный Институт» входит большое количество подключаемых модулей, которые могут быть легко включены в установки СУЭБ LibMeta, к примеру, для организации форума.

4.2 Профиль метаданных СУЭБ LibMeta

Одним из самых существенных недостатков многих схем метаданных ЭБ является то, что они рассматривают метаданные только в контексте описываемых ими данных и только как набор полей для поиска информации и индексирования ресурсов. В случае, когда ЭБ содержит один вид ресурсов, например, только книги, такой подход, возможно, является оправданным, однако современные ЭБ содержат разнообразные ресурсы, и в связи с этим такой подход неприемлем. Метаданные разных ресурсов должны содержать ссылки

друг на друга, но при этом, тем не менее, оставаться достаточно независимыми. Одним из наиболее удобных подходов к описанию такого рода метаданных является использование OWL-онтологий. Основной частью профиля метаданных в ЕНИП и «Научном институте» являются как раз такие онтологии. Общая схема профилей метаданных, применяемых в СУЭБ LibMeta, а также основных сущностей в данных профилях приведена на рис. 2.



Рис. 2. Профили метаданных СУЭБ LibMeta

Подробно профиль метаданных СУЭБ LibMeta был представлен на конференции в 2009 году [23]. Базовая часть профиля ЕНИП представлена в работе [1].

4.3 Интеграция СУЭБ LibMeta с другими информационными системами

Интеграция СУЭБ LibMeta с другими информационными системами осуществляется несколькими путями: во-первых, это интеграция в рамках ЕНИП, которая описана в разделе 4.1; во-вторых, это интеграция с системами подготовки публикаций; и, в-третьих, интеграция с универсальными агрегаторами.

4.3.1 Интеграция с системами подготовки публикаций

Для обеспечения интеграции с системами подготовки публикаций в СУЭБ LibMeta существует компонент загрузки и обновления данных и метаданных из внешних источников. Данный компонент ставит в соответствие каждому ресурсу, загружаемому в библиотеку извне, «состояние размещения», в зависимости от которого с ресурсом производятся действия по размещению и его периодическому обновлению. Также, когда ресурс уже размещён, возможно обновление ресурса по запросу от обновляющей (размещающей) стороны.

Описанная компонента настроена для работы с системой подготовки публикаций, применяемой в ЭБ «Научное наследие России», однако её архитектура рассчитана на простое добавление модулей для любых других систем подготовки публикаций. В плане развития СУЭБ LibMeta стоит разработка системы подготовки публикаций начального уровня, которая естественно будет интегрирована с СУЭБ LibMeta.

4.3.2 Интеграция с универсальными агрегаторами

Для интеграции с универсальными агрегаторами в СУЭБ LibMeta полностью реализованы стандарты OAI-PMH [19] и Dublin Core [8]. Кроме того, для интеграции с музейными системами существует отображение метаданных системы на концептуальную модель CIDOC-CRM [11]. Для использования других протоколов и форматов обмена данными и метаданными, в особенности, основанными на XML, не представляется сложной реализация модулей обмена данными.

В табл. 1 показано отображение содержательных метаданных СУЭБ LibMeta на Dublin Core и CIDOC-CRM. Отображение административных метаданных, фактически, тривиально, и поэтому не приводится. Также в таблице не приводится медиа-объект, поскольку в нём нет содержательных метаданных.

4.4 Требования СУЭБ LibMeta к аппаратно-программной платформе

Требования к аппаратной платформе при установке СУЭБ LibMeta варьируются в широких пределах в зависимости от выполняемых задач. Например, для ЭБ уровня школьной библиотеки требуется один x86 или x64 сервер с двумя процессорами, объёмом оперативной памяти 2 Гб и системой хранения данных объёмом 500 Гб. В случае более крупных ЭБ требуется наращивание в первую очередь системы хранения данных, и, во вторую – процессорной мощности.

Так как СУЭБ LibMeta реализована на кросс-платформенном языке Java, то она может быть установлена практически на любую операционную систему. Фактически единственным требованием к ОС является возможность установки среды Java и сервера приложений Apache Tomcat. В качестве СУБД может быть использована практически любая JDBC-совместимая СУБД, например, Microsoft SQL Server или PostgreSQL.

Следует отметить что установка СУЭБ LibMeta возможна на основе бесплатных ОС семейства linux в сочетании с бесплатной СУБД PostgreSQL, что позволяет снизить стоимость аппаратно-программной платформы до стоимости её аппаратной части.

Таблица 1. Отображение свойств метаданных СУЭБ LibMeta на Dublin Core и CIDOC-CRM

LibMeta	Dublin Core	CIDOC-CRM
Публикация		
Название	title	P102 Has title
Альтернативный заголовок	title	P102 Has title E35 Title
Аннотация	description	P3 Has note E62 String
Ключевые слова	subject	–
Источник	source	–
Авторские права	rights	P104 Subject of E30 Right
Web-адрес	–	–
Язык	language	P72 Has language E69 Language
Выпущен	date	–
Идентификатор	identifier	P1 Identified by E41 Appellation
Авторы	creator	P108 Produced by
Издатель	publisher	–
Редактор	contributor	–
Входит в состав	–	P46 Forms part of E18 Physical object
Включает	–	P46 Composed of E18 Physical object
Кол-во страниц	–	P57 Has number of parts E60 Number
Реферат	–	–
Библиографическое описание	escription	–
Полный код УДК	–	P2 Has type E55 Type
Примечания	–	P3 Has note E62 String
ББК	–	–
Основной код УДК	–	–
Персона		
Домашняя страница	–	–
Дата рождения	date	P97 Was born by E67 Birth
Адрес	–	P53 Has location E53 Place
Имя	title	P1 Identified by E41 Appellation
Пол	–	–
Фото	–	–
Учёная степень	–	–
Учёное звание	–	–
Дата смерти	date	P100 Died in E69 Death
Место рождения	–	P97 Was born by E67 Birth
Место смерти	–	P100 Died in E69 Death
Электронная почта	–	–
Телефон	–	–
Факс	–	–
Веб-страница	–	–
FTP-адрес	–	–
Музейный предмет		
Название	title	P102 Has title
Альтернативный заголовок	title	P102 Has title E35 Title
Аннотация	description	P3 Has note E62 String
Ключевые слова	subject	–
Источник	source	–
Держатель (место хранения)	–	P49 Has keeper E39 Actor
Состав	–	P46 Composed of E18 Physical object

LibMeta	Dublin Core	CIDOC-CRM
Автор описания	–	–
Состояние (сохранность)	–	P44 Has condition E3 Condition state
Количество предметов	–	P57 Has number of parts E60 Number
Номер	identifier	P1 Identified by E41 Appellation
Автор сбора	contibutor	P51 Has owner E39 Actor
Дата сбора	date	–
Дата поступления	date	–
География	–	P53 Has location E53 Place
Размеры	–	P43 has dimension E54 Dimension
Возраст	–	P92 Brought into existence E63 Beginning of existence
Способ поступления	–	–
Коллекция		
Название	title	P102 Has title
Тип коллекции	–	–
Ключевые слова	subject	–
Описание	description	P3 Has note E62 String
Количество элементов в коллекции	–	P57 Has number of parts E60 Number
Место хранения	–	P53 Has location E53 Place
Примечание	description	P3 Has note E62 String
Элементы коллекции	–	P46 Composed of E18 Physical object

4.5 Опыт внедрения СУЭБ LibMeta

Наиболее ярким примером внедрения СУЭБ LibMeta является проект ЭБ «Научное наследие России» [21], который разрабатывается в рамках одноимённой программы Президиума РАН с целью обеспечения сохранности и предоставления публичного доступа к научным трудам известных российских и зарубежных учёных и исследователей, работавших на территории России. Также некоторые из подсистем электронной библиотеки (системы хранения и представления электронных изданий конечным пользователям) создаются в рамках программы Президиума РАН «Информатизация». Общая координация и управление проектом осуществляются Межведомственным суперкомпьютерным центром РАН. Задачами подготовки электронных изданий и сопровождающей информации для размещения в хранилище данных электронной библиотеки занимаются ведущие библиотеки РАН, среди которых БАН, БЕН (Центральная библиотека и её отделения), ИНИОН.

Важной задачей ЭБ «Научное наследие России» являются интеграция существующих библиотечных ресурсов в ЕНИП РАН и обеспечение возможности централизованного доступа к ресурсам существующих хранилищ электронных изданий и метаданных об учёных и их научных трудах. Данная задача решается путём определения единой инфраструктуры распределённой системы, унификации форматов данных и протоколов взаимодействия компонентов системы, разработки единых регламентов подготовки и сопровождения электронных изданий.

5 Заключение

В настоящее время идёт становление ЭБ как в России, так и в мире. До сих пор не существует всеобъемлющего стандарта на ЭБ либо реализации ЭБ, удовлетворяющей всем существующим мировым стандартам. В статье рассмотрены основные требования, выдвигаемые к современным информационным системам и, в частности, ЭБ. Также приведён обзор наиболее распространённых стандартов, на которые следует ориентироваться при разработке ЭБ.

Созданная в ВЦ РАН ЭБ наиболее полно удовлетворяет приведенным выше требованиям, а построенная на её основе СУЭБ LibMeta позволяет научным институтам РАН, имеющим свои библиотечные, архивные или музейные фонды, создавать свои ЭБ и легко выставлять данные ресурсы в интернет как для научного, так и для широкого круга пользователей. На текущий момент работы по СУЭБ LibMeta практически завершены.

Литература

- [1] Бездушный А.Н., Бездушный А.А., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН. – М.: ВЦ РАН, 2006.
- [2] Berners-Lee T., Hendler J., Lassila J. The semantic web // Scientific Am. – 2001, May. – P. 3443.
- [3] Berners-Lee T., Shadbolt N., Hall W. The semantic web revisited// IEEE Intelligent Systems. – 2006, June.

- [4] DELOS an Association for Digital Libraries. – <http://www.delos.info>.
- [5] The DELOS Digital Library Reference Model – Foundations for Digital Libraries. Version 0.98 / L. Candela, D. Castelli, N. Ferro et al. – GEIE ERCIM, 2008.
- [6] Reference model for an open archival information system (OAIS): recommendation for space data system standards Blue book. CCSDS Secretariat, Program Integration Division (Code M-3). – National Aeronautics and Space Administration, 2002, January.
- [7] Consultative committee for space data systems. – <http://public.ccsds.org/default.aspx>.
- [8] Dublin core. – <http://dublincore.org/>.
- [9] The Dublin Core Metadata Element Set: an American national standard. – NISO Press, 2001.
- [10] CIDOC CRM Home page. – <http://cidoc.ics.forth.gr/index.html>.
- [11] Crofts N., Doerr M., Gill T., Stead S., Stiff M. Definition of the CIDOC Conceptual Reference Model. – January 2010.
- [12] The International Federation of Library Associations and Institutions. – <http://www.ia.org/>.
- [13] Functional Requirements for Bibliographic Records. – <http://www.ia.org/VII/s13/frbr/frbr.htm>.
- [14] Functional Requirements for Bibliographic Records Object-Oriented Definition and Mapping to FRBR_ER. – http://www.cidoc-crm.org/frbr_inro.html.
- [15] Publishing requirements for industry standard metadata. – <http://prismstandard.org>.
- [16] Library of Congress. Network Development and MARC Standards Office. MARC 21 concise formats. – Cataloging Distribution Service, Library of Congress, 2006.
- [17] Система форматов rusmarc. – <http://www.rba.ru/rusmarc/>.
- [18] Marc 21 xml schema. – <http://www.loc.gov/standards/marcxml/>.
- [19] Open archives initiative protocol for metadata harvesting. – <http://www.openarchives.org/pmh>.
- [20] Open archives initiative object reuse and exchange. – <http://www.openarchives.org/ore>.
- [21] Портал ЭБ «Научное Наследие России». – <http://sci-heritage.ras.ru/>.
- [22] Бездушный А.Н., Бездушный А.А., Нестеренко А.К. и др. Информационная Web-система «Научный институт на платформе ЕНИИП». – М.: ВЦ РАН, 2007.
- [23] Захаров А.А., Филиппов В.И. Логическая модель цифровых библиотек в онтологии ЕНИИП // «Электронные библиотеки: перспективные методы и технологии, электронные коллекции». Труды XI Всерос. науч. конф. RCDL'2009 (Петрозаводск, Россия, 17–21 сентября 2009 г.). – Петрозаводск: КарНЦ РАН, 2009. – Р. 487.

Digital library management system LibMeta

A.A. Zakharov, V.A. Serebryakov

The article presents digital library management system (DLMS) LibMeta which is dedicated for science institutes in RAS. Its architecture is described along with standards which form its foundation. Also digital library "Russian scientific Heritage" based on DLMS libMeta is presented.