

Ю.И. Шокин, А. М. Федотов, О.Л. Жижимов, О. А. Федотова
Институт вычислительных технологий СО РАН,
Новосибирский государственный университет,
Государственная публичная научная библиотека СО РАН

Система управления электронными библиотеками¹

Аннотация: Начиная с 1998 года в ИВТ СО РАН, ведутся работы по созданию ИРИС СО РАН (Интегрированной Распределённой Информационной Системы), организованной в виде электронной библиотеки. Статья посвящена истории создания и описанию технологических решений, применяемых при создании системы. Описываются архитектура информационной системы и принципы интеграции с внешними источниками, правила представления и преобразования метаданных, а также описана работа со словарями, которые используются для систематизации и классификации информационных ресурсов, и моделированию связей между ними.

Ключевые слова: ИРИС, информационная система, электронная библиотека, словарь-справочник, классификация информационных ресурсов, цифровой репозиторий, информационно-поисковый тезаурус, ключевые термины, протокол OAI-PMH, метаданные.

Введение

В начале 1998 г. в Сибирском отделении РАН была сформирована целевая программа развития информационных ресурсов Отделения под общим названием “Электронная библиотека Сибирского отделения РАН”. Для решения проблемы информационной обеспеченности сотрудников Отделения было принято решение о создании собственной универсальной Интегрированной Распределённой Информационной Системы (ИРИС) Отделения [1,2], содержащей полнофункциональную систему об интеллектуальном потенциале Отделения (информационную систему об институтах, сотрудниках, достижениях и др. аспектах, связанных с работой Отделения) и систему электронной поддержки сбора и накопления научной информации (электронных атласов, электронных коллекций, баз данных и т.п.).

Основные направления программы связаны с формированием собственных электронных ресурсов по основным отраслям наук (математика, науки о земле, химия, биология, археология и др.), созданию и поддержке электронных

¹ Работа выполнена при частичной поддержке РФФИ (проекты 12-07-00472, 13-07-00258), президентской программы «Ведущие научные школы РФ» (грант НШ–5006.2014.9).

коллекций и электронных публикаций, организации удобных систем доступа к библиотечным и библиографическим базам данных и базам данных Институтов Отделения [3]. Результаты работ должны обеспечить:

- о Единую распределенную информационную среду Отделения.
- о Информационную поддержку исследований по фундаментальным и прикладным направлениям.
- о Поддержку профессионально-ориентированных систем подготовки и обмена научных документов.
- о Поддержку профессионально-ориентированных систем доступа и интерфейсов с хранилищами данных.
- о Коллективное использование приобретаемой электронной литературы, каталогов, баз данных и библиографических изданий.
- о Поддержку электронных версий научных журналов, издаваемых институтами Отделения.
- о Поддержку принятия и реализации организационных и управленческих решений в Отделении.

Основные ресурсы, созданные тогда не потеряли своей актуальности до сих пор. В первую очередь здесь стоит упомянуть электронный атлас «Биоразнообразие животного и растительного мира Сибири» [4], систему поддержки проведения конференций [5], виртуальной музей СО РАН [6], база данных организаций и сотрудников СО РАН [7]. В результате работ по программе сложилось понимание того, что информационная система для поддержки научных исследований ИРИС должна основываться на использовании концепции электронных (цифровых) библиотек [8-11]. В рамках нашего подхода цифровые библиотеки рассматриваются как отдельная конкретная технология работы с цифровой информацией, образующая новый класс информационных систем (ИС), предназначенных для управления информационными ресурсами [12, 13].

Под термином *электронная библиотека* (ЭБ) в данной работе будем понимать систему управления структурированными каталогизированными коллекциями разнородных электронных (цифровых) объектов (ресурсов). ЭБ не только обеспечивает многосторонний поиск и навигацию по каталогам (в отличие от печатных изданий, микрофильмов и других носителей), но и предоставляет пользователю непосредственно найденный ресурс (публикацию, документ, фотографию, описание факта и др.), а также дополнительные сведения о нем, например, географическую привязку, информацию об авторах, информацию о фактах, библиографию, перечень организаций и т. д.

Концепция ИРИС

На сегодняшний день наиболее эффективным способом решения проблем организации доступа к распределенным информационным ресурсам является организация информации о них в информационные системы, обличенных в форму электронных библиотек². В работах [14,15] были сформулированы основные принципы реализации ИРИС, основанные использовании идеи электронных библиотек.

Отметим, что ЭБ – явление относительно новое, но уже достаточно популярное [16]. Тем не менее ЭБ сегодня следует рассматривать как множество слабосвязанных сущностей, объединяемых на первый взгляд только общим названием [13, 17]. Современное название «электронная библиотека» – это не только и не столько дань моде, сколько попытка охарактеризовать новый феномен – возникновение принципиально нового класса систем, призванных аккумулировать и распространять информацию в электронной форме [18]. А большой интерес к самим системам данного класса объясняется потребностями общества и ростом возможностей по их удовлетворению [19].

В связи с этим можно сформулировать основные цели, стоящие перед ЭБ (системами управления информационными ресурсами) [20]:

² По-простому, понятие электронной библиотеки заключается в том, что любой ресурс должен быть *стандартным* образом каталогизирован, снабжен *метаданными*, правилами доступа и уникальным идентификатором.

- управление информационными ресурсами;
- обеспечение и управление доступом к информации;
- долговременное хранение информации;
- сохранение научного и культурного наследия;
- поддержка аналитической работы с информацией;
- повышение эффективности научных исследований и обучения.

В существующих разработках ЭБ, как правило, поиск и доступ к информации обеспечиваются только посредством визуальных графических интерфейсов. Это хорошо для пользователя-человека, но не годится для пользователя-системы. Для реализации функций поиска вне графических интерфейсов требуется поддержка специальных сетевых сервисов и языков запросов. В идеальном случае все ИС должны поддерживать единый поисковый профиль и единый язык запросов [12].

Однако в общем случае под словосочетанием «электронная библиотека» могут фигурировать совершенно различные объекты, такие как архивы цифрового контента и наборы программного обеспечения для управления этим контентом. ЭБ может называться система сетевых сервисов, предоставляющих доступ к цифровому контенту и объединенных единой системой управления этим доступом [20]. Такое определение ЭБ полностью соответствует определению традиционной библиотеки как организации в системе, например, Министерства культуры [12].

Ввиду того, что информация в ИС отображает некоторые сущности реального мира (физические объекты: предметы, процессы, явления, персоны, публикации, документы, алгоритмы, программы, файлы, факты, ключевые термины и т. д.), следует рассматривать ИС как множество информационных объектов – наборов данных, представляющих (описывающих) эти сущности в ИС. Отметим, что разработка модели ЭБ должна использовать онтологические описания и концептуальные модели, обобщающие накопленный опыт в сфере создания и использования ЭБ [20]. Неплохой обзор существующих концептуальных моделей ЭБ приведен в [22].

Онтологическая модель ЭБ ИРИС основана на концептуальных моделях ЭБ RM OAIS³ [23] и DELOS DLRM⁴ [24].

В соответствии с концептуальной моделью DELOS Информационный Ресурс (ИР) – это абстрактное понятие, выражаемое экземплярами одной из своих специализаций. В частности, экземплярами понятия ИР являются экземпляры информационного объекта любого типа (например, документы, базы данных, коллекции, функции и т.п.). Каждый ресурс в соответствии с моделью DELOS:

- о имеет идентификатор;
- о организован в соответствии с описанием ресурса. Ресурс может быть сложным и структурированным, поскольку он, в свою очередь, может состоять из меньших ресурсов и иметь связи с другими ресурсами;
- о может регулироваться функциями, управляющими его жизненным циклом;
- о выражается через информационный объект;
- о должен быть описан метаданными, а также может быть описан или дополнен дополнительными метаданными и аннотациями.

В основе реализации Системы Управления Электронными Библиотеками (СУЭБ) в ИРИС лежит метамодель, исходящая из того, что каждый информационный ресурс характеризуется набором присущих ему атрибутов, и методов, характеризующих его свойства и связи с другими ресурсами. Эффективным средством описания информационных объектов являются метаданные – данные, являющиеся неотъемлемой частью информационного объекта и описывающие реальный объект или группу объектов.

Каждый информационный объект в ИС состоит из (см. рис. 1):

³ Open Archival Information System

⁴ Digital Library Reference Model

- о Информационного содержания объекта (первичный информационный объект: например, изображение, полный текст и т.д.) – объект, который может использоваться самостоятельно;
- о объект метаданные – объект, главная цель которого состоит в том, чтобы дать информацию об ИР (как правило о первичном информационном объекте);
- о объект аннотация – объект, главная цель которого состоит в том, чтобы аннотировать ИР или его часть. Примеры таких аннотаций включают примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать ИР, содержат либо поддержку, либо детальные объяснения, либо информацию о том, как можно использовать ИР.

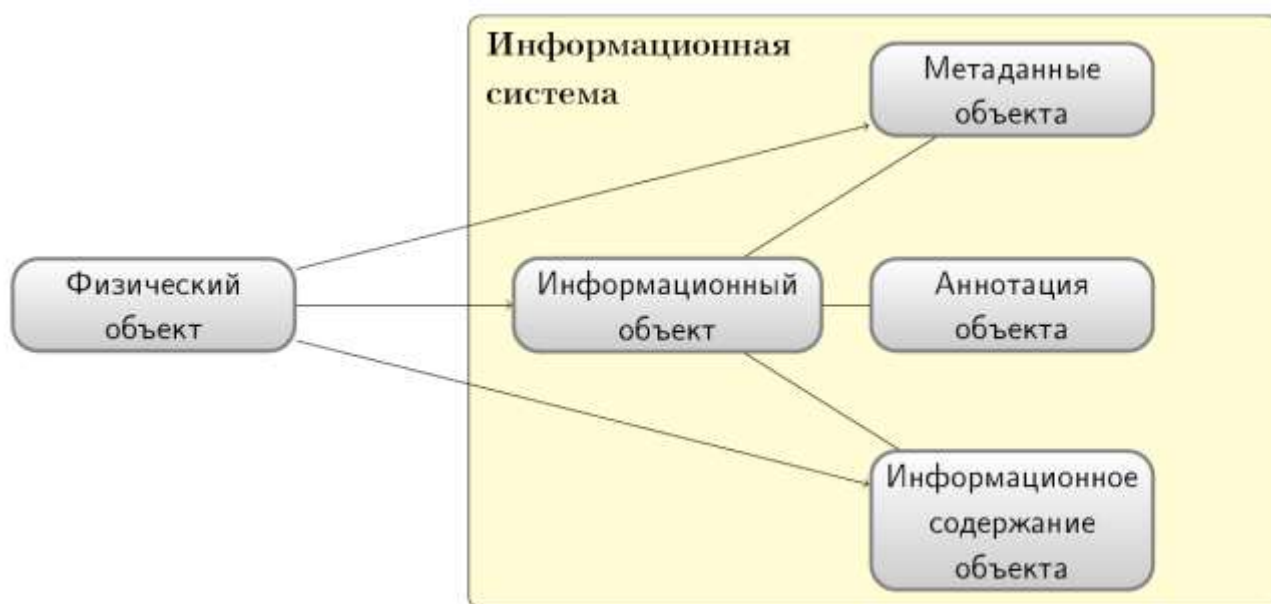


Рис.1. Структура сущностей ИС

Документы

В информационном пространстве все информационные ресурсы: события, факты, программы и любые другие сущности реального или виртуального мира существуют только в форме некоторых информационных объектов [25, 26], концептуальная модель которых была описана выше.

Бельгийский ученый П. Отле – пионер и основоположник науки «информатика» – в своем трактате о Документации [27] определяет расширение понятия (термина) Документ⁵: «материальный объект, содержащий информацию, специально предназначенный для ее передачи в пространстве и времени» – который трактуется как основной «объект», с которым оперирует любая информационная система [28]. Таким образом, Документ — это информационный объект, представляющий структурированное описание реальной сущности (объекта, субъекта, факта или понятия), совокупность которых составляют информационное наполнение системы.

В ИС Документ – это целостный информационный объект, представленный в цифровом виде, имеющий некоторый стандартный набор атрибутов и функций и допускающий однозначную идентификацию. Документ может описывать статью из журнала, сам журнал, персону, оцифрованное изображение, экспериментальные данные, программа или вычислительный алгоритм, база данных, фрагмент базы данных и т.п. [25]. Однако документ не обязан содержать полный текст статьи, но в качестве атрибута он может содержать указатель на хранилище данных (репозиторий), где полный текст хранится. Кратко требования к документу можно охарактеризовать так:

- документ имеет уникальный идентификатор;
- документ имеет структурированное описание (метаданные);
- документ имеет набор атрибутов (*свойств*) и методов (функций);
- взаимодействие с документом (например, работа с атрибутами) происходит через набор *методов*.

Ключевым моментом в работе с документами является использование метаданных для формирования структуры, схемы данных и свойств документов – ин-

⁵ от лат. Documentum — доказательство

формационных элементов системы и ведения каталога системы. Любая ЭБ опирается на метаинформацию, содержащую онтологию, описывающую принципы организации информации. Онтология, описывающая конкретную предметную область, конкретизируется в схеме данных (атрибутов), описывающих информацию в метаданных [26].

Электронные библиотеки

Как мы уже отмечали, что электронные библиотеки – это распределённые каталогизированные информационные системы, позволяющие хранить, обрабатывать, распространять, анализировать, а также организовывать поиск в разнообразных коллекциях электронных (цифровых) документов. Основная задача, решаемая электронными библиотеками – это управление информационными ресурсами и «интеграция информационных ресурсов (включая поддержку унифицированного доступа к ним), а также эффективная навигация в них» [25].

Под интеграцией информационных ресурсов понимается их объединение с целью использования (с помощью удобных и унифицированных пользовательских интерфейсов) разнородной информации с сохранением ее свойств, особенностей представления и пользовательских возможностей манипулирования с ней. При этом объединение ресурсов не обязательно должно осуществляться физически, оно может быть виртуальным, главное — оно должно обеспечивать пользователю восприятие доступной информации как единого информационного пространства. В частности, такие системы обеспечивают работу с гетерогенными наборами и базами данных или системами баз данных, обеспечивая пользователю эффективность информационных поисков независимо от особенностей конкретных систем хранения ресурсов, к которым осуществляется доступ [26].

Под эффективной навигацией в информационной системе понимается возможность для пользователя находить интересующую его информацию с наибольшей полнотой и точностью при наименьших затратах усилий во всем доступном

информационном пространстве. При таком подходе хорошо известные информационно-поисковые системы, используемые в информационных системах и базах данных, являются частными случаями навигационных средств [25, 26].

В настоящее время существуют достаточно мощные ИС, удовлетворяющие в той или иной степени потребности научных работников в информации, однако основной недостаток большинства систем – ограниченность интеграции ресурсов как внутри каждой из них, так и с внешними системами. Основу разработки ЭБ составляют стандарты и международные рекомендации, формирующие профиль ЭБ, под которым понимается один или набор нескольких базовых нормативно-технических документов (стандартов и спецификаций), ориентированных на решение определенной задачи (реализацию заданной функции либо группы функций приложения или среды), с указанием, если нужно, выбранных классов, подмножеств, опций базовых стандартов, необходимых для выполнения конкретной функции [29]. Наиболее важным является профиль метаданных информации, циркулирующей в системе. Выбор профиля метаданных должен основываться на выполнении следующих требований [20, 25, 30, 31]:

- включать описания основных типов информации, требующейся для поддержки научно-образовательной деятельности;
- быть открытым, т. е. обеспечивать доступ к информации в соответствии с ее описанием (метаданными);
- быть расширяемым, т. е. обеспечивать возможность детализации описаний;
- обеспечивать возможность интеграции информации;
- обеспечивать возможность уникальной идентификации информации;
- обеспечивать отбор, систематизацию и классификацию информации;
- обеспечивать возможности размещения и поиска информации в распределенной среде;

- быть ориентированным на современные технологии описания и использования информации;
- обеспечивать возможность интероперабельности с другими системами.

Отметим, что серьезной проблемой является идентификация информационных ресурсов [32], позволяющая получать библиографические сведения, а также устанавливать связи определенного ресурса с другими фактами и объектами.

При работе с цифровыми объектами уже выработан определенный набор стереотипов, отсутствие которых вызывает дискомфорт [12]. Одним из элементов этого набора является наличие взаимных ссылок между цифровыми объектами, например, в виде гиперсвязей в пользовательских графических интерфейсах просмотра информации. Реализация взаимных ссылок в цифровых документах не представляет большой сложности, однако здесь прослеживается своя специфика. Во-первых, электронный объект с реализованными связями уже не совсем соответствует своему печатному оригиналу. Во-вторых, внедренные в объект связи должны быть гарантированно актуальными. Так появляется требование обеспечения ссылочной целостности данных. Это очень жесткое требование, которое тяжело соблюсти даже в хорошо формализованных системах управления БД. Приемлемым решением здесь может быть замена жестких гиперссылок динамическими связями между документами на уровне системы управления. Результат – новый цифровой объект как самосогласованное хранилище цифрового контента или БД цифровых объектов.

Существует достаточно много технологических разработок ИС для ЭБ, так или иначе ориентированных на поддержку научных исследований, например,

euroCRIS⁶, eLibrary⁷, Информика⁸, MathNET⁹. В большей степени удовлетворяет информационным потребностям научно-образовательного сообщества в информации система ИСИР (ЕНИП) РАН [30, 33].

Наиболее важным выводом из вышесказанного является следующий: информационная модель ЭБ должна быть многоуровневой и состоять как минимум из нескольких компонент [21, 25, 26]: хранилища данных (репозиторий), сервера метаданных, сервера приложений (диспетчера), словарей-справочников (см. рис.2).

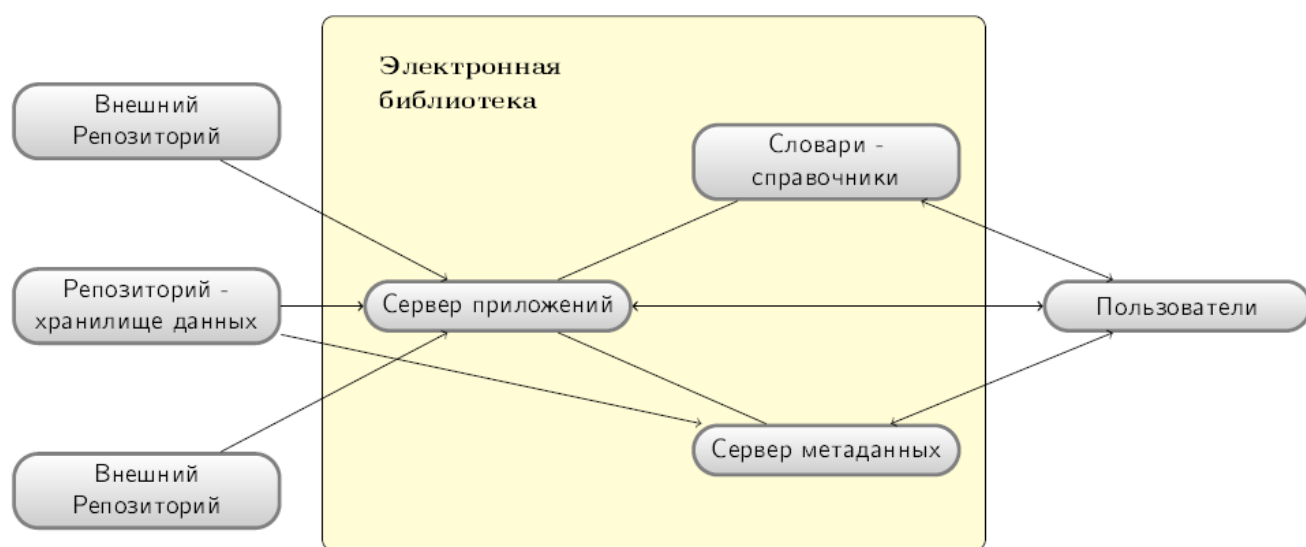


Рис.2. Архитектура ЭБ

Цифровые репозитории

Для организации системы долговременного хранения информационных ресурсов (репозиториях цифровых объектов) международной организацией по стандартизации (ISO) предложен стандарт ISO-14721 (Open Archive Information System – OAIS¹⁰) [23]. Эталонная модель для стандарта OAIS – это концептуальная модель,

⁶ <http://www.eurocris.org/>

⁷ <http://elibrary.ru/>

⁸ <http://www.informika.ru/>

⁹ <http://www.mathnet.ru/>

¹⁰ Открытая Архивная Информационная Система.

основанная на расширенной схеме данных Dublin Core [34]. Эта модель была использована многими организациями для разработки наборов метаданных и создания крупных хранилищ цифровых объектов.

На основе данной модели создана концепция «институционального репозитория» как системы долговременного хранения, накопления информации и обеспечения надежного доступа к цифровым объектам, представляющим собой результат интеллектуальной деятельности научного или образовательного учреждения. К основным особенностям институционального репозитория относятся:

- обеспечение разграниченного доступа к разнородным цифровым объектам (публикациям, изображениям и т. д.);
- организация доступа к информационным ресурсам для мирового сообщества (в том числе с помощью полнотекстового индексирования мировыми поисковыми системами);
- унифицированный доступ к метаданным по стандартным протоколам (поддержка интероперабельности);
- возможность организации единой точки доступа к информационным ресурсам;
- сохранение других информационных ресурсов, в том числе неопубликованных, таких как диссертации, препринты и технические отчеты, программное обеспечение, мультимедиа и т. д.

Согласно данным сайта OpenDOAR¹¹, большинство институциональных репозиторий основано на свободном программном обеспечении и построено в рамках модели OAIS на базе технологий открытых систем. В настоящий момент в мире насчитывается более десятка систем поддержки институциональных репозиторий, наиболее популярные из них DSpace [35] (свыше 41% установок), E-Prints [36], Fedora [37] (сравнительную характеристику этих систем и описание используемых

¹¹ The Directory of Open Access Repositories – http://www.open_doar.org/

в них информационных моделей можно найти в [38]). Процесс интеграции репозитория в среду ЭБ для этих систем отличается лишь несущественными деталями и основан на модели агрегирования и распространения метаданных. Применение этой модели закреплено в протоколе OAI Protocol for Metadata Harvesting (далее OAI или OAI-PMH) [39], который поддерживается большинством систем, предназначенных для хранения информационных ресурсов.

В качестве репозитория для ИС нами была выбрана система DSpace. Выбор обусловлен тем, что она является самой популярной в мире и уже эксплуатируется в СО РАН¹² (а так же в ряде других институтов и университетов России) на протяжении десяти лет.

Кроме того, система DSpace имеет ряд привлекательных особенностей. Для базовой организации данных зафиксирована определенная модель данных (DIM – внутренний формат данных DSpace), основанная на схеме Dublin Core и ее расширениях. При некотором напряжении при помощи этой схемы можно отобразить основные элементы всех используемых в настоящий момент форматов. Система с помощью фильтров метаданных, которые используются для преобразования метаданных из внутренней схемы в схемы, пригодные для обмена метаданными с внешними системами на основе XSLT¹³ трансформаций, позволяет конвертировать и индексировать метаданные в разнообразных форматах (МЕКОФ¹⁴, MODS¹⁵, METS¹⁶, QDC¹⁷, MARC¹⁸ и др.). Система хранит информацию о пользователях, поддерживает авторизацию и разграничивает доступ к содержимому репозитория по группам, сетевым адресам и на основе протокола LDAP [40], что при создании ИС дает возможность использовать уже существующую систему аутентификации поль-

¹² <http://elib.nsc.ru:8080/jspui/>

¹³ <http://www.w3.org/TR/xslt/>

¹⁴ Этот формат наиболее привлекателен в России, т. к. поддержан ГОСТом [35].

¹⁵ Metadata Object Description Schema (MODS) – <http://www.loc.gov/standards/mods/>

¹⁶ Metadata Encoding and Transmission Standard (METS) – <http://www.loc.gov/METS>

¹⁷ Qualified Dublin Core (QDC) [29].

¹⁸ Machine-Readable Cataloging (MARC) – <http://www.loc.gov/marc/>

зователей (а не разрабатывать свою собственную) и достаточно легко дифференцировать публичные и служебные ресурсы, оставляя при этом свободный доступ к метаданным. Провайдеры данных для протоколов OAI-PMH [39], Z39.50 [41, 42] и SRW/SRU [43], стандартных для библиотечного сообщества, позволяют разрабатывать программный интерфейс для взаимодействия ИС поддержки ЭБ с хранилищем данных, построенном на основе DSpace.

Отметим, что для обеспечения работы ЭБ в рамках сформулированных выше функциональных требований, недостаточно метаданных, используемых для долговременного сохранения цифровых объектов (нужны дополнительные метаданные для организации связей между документами, способов представления документов и т. д., что подробнее изложено в следующем разделе), поэтому необходимо, чтобы метаданные хранились в системе поддержки ЭБ, независимой от той, которая использовалась при их создании.

Для синхронизации метаданных информационных ресурсов между ИС ЭБ и хранилищем данных используется сервис, основанный на использовании протоколов OAI-PMH [39], SRW/SRU [43] и сервера приложений на основе ZooPARK-ZS [44] (в зависимости от типа хранилища данных). В задачи этого сервиса входит извлечение метаданных из репозитория, конвертирование (при необходимости) и передача их серверу метаданных. Под конвертированием здесь понимается как преобразование метаданных и схем метаданных (например, QDC [34] в ГОСТ 7.19 [45] или RUSMARC [46]), так и преобразование форматов (например, в XML [47] или ISO-2709 [48]).

Следует подчеркнуть, что при взаимодействии с удаленными системами, обмен метаданными должен происходить согласованно, с использованием форматов обмена.

Выбор метаданных для ЭБ

Для поддержки сложных функций поиска и классификации необходимо обеспечить возможность поиска по атрибутам, полнотекстового поиска, а также просмотра ресурсов по категориям и словарям-классификаторам.

В существующих ИС информационные ресурсы разрознены, недостаточно систематизированы и структурированы. В ходе создания их описаний недостаточное внимание уделяется вопросам интероперабельности: слабо применяются соглашения и рекомендации по стандартизации представления документов и средства интеграции разнородных информационных ресурсов. Под интероперабельностью ИС понимается степень ее способности взаимодействовать с другими ИС, в том числе и с человеком. Но если при взаимодействии с человеком (как с ИС) основная нагрузка на достижение взаимопонимания ложится на последнего, способного обработать даже плохо организованную информацию, то для обеспечения эффективного взаимодействия между собственно ИС требуются специальные технологические методы и общие соглашения. Это влечет за собой необходимость соблюдения соответствия всех схем данных, интерфейсов и протоколов международным стандартам и рекомендациям [12, 20].

В работах [20, 21] был определен профиль ЭБ как необходимый набор стандартов и компонентов информационной системы, ориентированной на научные исследования.

В настоящее время существует большое количество систем метаданных, предназначенных для описания различных классов информационных объектов. Использование систем метаданных (схем данных) пока еще недостаточно формализовано. ИС, ориентированные на одинаковые классы информационных объектов используют различные, зачастую оригинальные системы метаданных и фор-

маты метаописаний, а также разные подходы к решению прикладных задач. Устранением подобных несоответствий занимаются многие организации по всему миру, например, такие как W3C¹⁹, DCMI[29], OCLC²⁰, IFLA²¹, IETF²², ISO²³.

Метаданные необходимы для решения следующих задач:

- предоставление сведений об объекте для получения представления о его содержании, структуре, способах использования и т. д.;
- сбор и систематизация информации об объектах описания;
- выбор из множества объектов определенного подмножества по формальным признакам и сопоставление объектов по формальным признакам;
- внутрисистемные технологические задачи, связанные с обеспечением подготовки объектов, размещением объектов в информационном фонде и т. д.;
- внешние технологические задачи, связанные, прежде всего, с обменом данными с внешними информационными системами.

Основу содержания ЭБ в ИРИС составляют документы (информационные объекты), представляющие основные типы сущностей:

- субъекты: персона, организация и т. д.;
- объекты (единицы хранения): публикация, документ, факт, научный результат, мероприятие, фотография и др.;
- отношения: понятие, ключевой термин, событие, время, место и т.п.

В отличие от общепринятых ЭБ указание на субъекты дается с помощью ссылки на экземпляр сущности «субъект», что позволяет корректно решать задачу идентификации объектов.

¹⁹ World Wide Web Consortium (W3C) – <http://www.w3.org/>

²⁰ Online Computer Library Center – <http://oclc.org/>

²¹ International Federation of Library Associations – <http://www.ifla.org/>

²² Internet Engineering Task Force (IETF) – <http://www.ietf.org/>

²³ International Organization for Standardization – <http://www.iso.org/>

Используемый профиль определяет список элементов данных (полей), необходимых для создания записи соответствующего типа, и раскрывает содержание элементов данных. Для эффективной работы сервера приложений используется набор словарей-классификаторов, содержащих как классификационные признаки, так и наборы ключевых терминов (с отношениями порядка), по которым производится систематизация и классификация материала.

Для формирования метаданных применяются несколько стандартов, являющихся расширениями рекомендаций Dublin Core²⁴ и Qualified Dublin Core (QDC). Для документов нами была расширена стандартная схема метаданных QDC полями, включающими основные требования государственного стандарта МЕКОФ [45].

Словари (ключевые признаки, ключевые термины) – это особый вид метаданных, отражающих наиболее существенные свойства объекта и имеющие наиболее важное значение с точки зрения ИС. Специфика словарей определяется терминологией конкретной предметной области, которой посвящена ЭБ. Необходимо рассматривать различные типы ключевых терминов (ключевые термины в стандартном понимании; ключевые термины, описывающие персону; ключевые термины, описывающие организацию; ключевые термины, описывающие временные периоды; ключевые термины, описывающие географические понятия), а также тематические словари-классификаторы, тезаурусы, описания предметной области данной научной школы, и классификаторы документов в соответствии с МЕКОФ.

Существует ряд российских (например, УДК²⁵, ГРНТИ²⁶) и международных (например, MSC-2000²⁷, ORTELIUS²⁸) словарей для классификации научных данных.

²⁴ <http://www.dublincore.org/>

²⁵ УДК – Универсальная десятичная классификация.

²⁶ ГРНТИ – Государственный рубрикатор научно-технической информации.

²⁷ MSC-2000 – Математический классификатор – <http://www.ams.org/msc/msc.html>

²⁸ The «Ortelius Thesaurus on Higher Education» – http://cordis.europa.eu/cerif/src/sum_concl.htm

Однако в целом эти словари содержат только общенаучную информацию и не подходят (хотя использовать их все равно нужно) для систематизации материалов.

Метаданные существенным образом зависят от природы и структуры объектов реального мира, от способа представления их в виде информационных объектов и от специфики ИС. Учитывая это, необходимо классифицировать описываемые объекты. Совокупность правил, достаточная для формирования метаданных в определенном классе ИС и (или) для решения определенного ряда задач над информационными объектами, представляет собой систему метаданных.

Практическая реализация

Рассмотренная модель ЭБ реализована в виде Системы Управления ЭБ (СУЭБ ИРИС), созданной и эксплуатируемой в ИВТ СО РАН с 2004 года [25]. СУЭБ ИРИС оперирует электронными коллекциями документов. Электронная коллекция – это набор документов, объединённых по смысловому признаку и имеющих одинаковую структуру (схему данных) [26]. СУЭБ позволяет работать с двумя видами коллекций: каталогами и тезаурусами. Принципиальное отличие каталогов от тезаурусов состоит в том, что в тезаурусах можно организовывать иерархические зависимости (родитель-потомок, часть-целое и т.п.) между записями. Коллекции-каталоги предназначены для хранения и обработки метаданных о документах различной природы: публикации, ключевые термины, персоны, организации, фотографии и т. д. Коллекции-тезаурусы предназначены для работы со словарями-классификаторами.

Сервер метаданных СУЭБ содержит служебную коллекцию Основной каталог метаданных, которая содержит документы, описывающие все метаданные, которые можно использовать в системе. Документы Основного каталога содержат описания схемы метаданных QDC, расширенной метаданными для соответствия МЕКОФ и описания служебных метаданных, описывающими структуру объектов, пользовательские интерфейсы, ассоциативные связи между документами, права

доступа к документам и т. д. (при желании он может быть расширен новыми метаданными).

Априори каждая коллекция (в зависимости от вида) имеет минимальный обязательный набор метаданных. Администратор коллекции имеет возможность доопределить схему метаданных коллекции, исходя из имеющихся метаданных из основного каталога.

В СУЭБ представлено два вида ассоциативных связей между документами (записями): жесткие и мягкие. Жесткие связи реализованы средствами СУБД путем ссылок на первичные ключи записи. К сожалению, такой тип связи не защищен от нарушения целостности (в случае неправильного изменения или удаления записи). Мягкие связи реализуются через процедуру поиска соответствий. Такой способ установления связей защищен от любых нарушений целостности БД и достаточно удобен пользователям, поскольку для указания на необходимость связи используются наглядные мнемонические определения. Соответствия устанавливаются следующими способами:

- Ссылка на идентификатор записи – уникальный, в пределах одной коллекции, текстовый код, формируемый в рамках конкретной коллекции по определенным правилам. Например, для коллекции, содержащей описания персон, идентификатор формируется (на русском языке) последовательно из фамилии – инициалов – года рождения. Отметим, что за десять лет эксплуатации СУЭБ не было зафиксировано ни одного конфликта при формировании идентификаторов.
- Ссылка на ключевой термин – особый вид метаданных, выбираемый из словаря ключевых терминов, по существу представляющий собой тезаурус предметной области коллекции. Ссылка определяет запись, в которой данный ключевой термин присутствует в метаданных.

Для внутреннего долговременного хранения цифровых объектов используется репозиторий DSpace [35]. Стандартная схема метаданных DSpace, основанная

на схеме DСMI [34], расширена полями, отвечающими основным требованиям МЕКОФ [45]. Для поддержки процесса наполнения полнотекстовых БД созданные профили метаданных зарегистрированы в системе DSpace, и в соответствии с ними настроены рабочие процессы, а также пользовательский интерфейс системы.

С целью организации обмена метаданными между DSpace и сервером метаданных (а также с другими системами с расширенным профилем) создан специальный сервис, выполняющий преобразование метаданных из внутренней схемы DSpace в другие схемы метаданных, в том числе и в схему DСMI, с использованием квалификаторов (QDC²⁹), а также в схему МЕКОФ (представление ISO2709 или XML). Реализован OAI-PMH сервис, который в пакетном режиме периодически, в соответствии с расписанием, проводит синхронизацию метаданных репозитория и сервера метаданных. Для заполнения основного каталога метаданных в соответствии с созданными схемами метаданных используются контролируемые словари из справочного блока сопровождения. Для обеспечения интероперабельности данных DSpace также задействован сервер приложений на основе ZooPARK-ZS [44], реализующий доступ к метаданным системы по протоколам Z39.50 [41, 42] и SRW/SRU [43].

Разработанная модель может быть использована как типовая модель системы для работы с документами, связанными с научно-образовательной деятельностью, поскольку решает основные задачи, предъявляемые к этим системам: обеспечение системы надежного долговременного хранения цифровых (электронных) документов с сохранением всех смысловых и функциональных характеристик исходных документов; обеспечение «прозрачного» поиска и доступа пользователей к документам, как для ознакомления, так и для анализа содержащихся в них фактов; организация сбора информации по удаленным цифровым репозиториям, поддерживающим протоколы OAI-PMH, SRW/SRU, Z39.50.

²⁹ Qualified Dublin Core (QDC) – <http://www.dublincore.org/documents/dcmi-terms/>

Рассмотренная модель ЭБ реализована на примере научной школы Алексея Андреевича Ляпунова – основателя теоретического программирования и российской кибернетики, в виде ЭБ по моделям динамики биосферы, а также в виде ЭБ учебных пособий³⁰ по курсам «Современные проблемы информатики и вычислительной техники», «Вычислительные системы», «Информатика» и «Экология» и др.

Литература

1. Шокин Ю.И., Федотов А.М. Распределенные информационные системы // Вычислительные технологии. - 1998. - Т.3. - № 5. - С.79-93.
2. Шокин Ю.И., Федотов А.М. Электронная библиотека Сибирского отделения РАН // Электронные библиотеки: рос. науч. электронный журн. – 1999 - Том 2 - Выпуск 4. [Электронный ресурс].
<http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part4/fedotov> (дата обращения: 04.09.2014).
3. Шокин Ю.И., Федотов А.М., Жижимов О.Л., Мазов Н.А. Интегрированная распределенная информационная система (ИРИС) Сибирского отделения РАН // В сб.: Материалы выездного заседания научно-координационного совета по целевой программе "Информационно-телекоммуникационные ресурсы СО РАН", Иркутск, 29-30 августа 2002: Ин-т географии СО РАН, 2003, с. 139-149.
4. Федотов А.М., Артемов И.А., Ермаков Н.Б., Красников А.А., Потемкин О.Н., Рябко Б.Я., Федотов А.А., Хорев А.Г. Электронный атлас «Биоразнообразие растительного мира Сибири» // Вычислительные технологии. - 1998. - Т.3. - № 5. - С.68-78.
5. Федотов А.М., Гуськов А.Е., Молородов Ю.И. Информационная система поддержки проведения конференций СО РАН // В сб.: Материалы выездного заседания научно-координационного совета по целевой программе "Информационно-телекоммуникационные ресурсы СО РАН", Иркутск, 29-30 августа 2002: Ин-т географии СО РАН, 2003, с. 91-110.
6. Шокин Ю.И., Ламин В.А., Федотов А.М. и др. Виртуальный музей Науки и Техники СО РАН// В сб.: Материалы выездного заседания научно-координационного совета по целевой программе "Информационно-телекоммуникационные ресурсы СО РАН", Иркутск, 29-30

³⁰ <http://fedotov.nsu.ru/lecture.php>

- августа 2002: Ин-т географии СО РАН, 2003, с. 118-125.
7. Леонова Ю.В., Клименко О.А., Федотов А.М. Информационная система «База данных организаций и сотрудников СО РАН». - Новосибирск: РИЦ Прайс-Курьер. - 2005. – 55 с.
 8. Федотов А.М., Шокин Ю.И. Электронная библиотека Сибирского отделения РАН // Информационное общество. - 2000. - № 2. - С.22-31.
 9. Шокин Ю.И., Федотов А.М. Библиотека, работающая круглосуточно // ЭКО. - 2000. - № 6. - С.163-172.
 10. Шокин Ю.И., Федотов А.М. Информационная система Сибирского отделения РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Сб. докладов. Протвино, 2000. - С.6-15.
 11. Шокин Ю.И., Федотов А.М. Электронная библиотека Сибирского отделения РАН // Информационно-библиотечное обеспечение науки. Проблемы интеграции информационных ресурсов: Сб. науч. тр. - 2000. - М. - С.118-128.
 12. Жижимов О. Л., Мазов Н.А., Федотов А.М. Некоторые заметки об эволюции цифровых репозиториях традиционных библиотек к полнофункциональным электронным библиотекам // Вестник Владивостокского государственного университета экономики и сервиса. Территория новых возможностей. 2010. Т. 7. № 3. С. 55-63.
 13. Антопольский А.Б., Вигурский К.В. Концепция электронных библиотек [Электронный ресурс]. Электронные библиотеки: рос. науч. электронный журн. Т. 2. Вып. 2. 1999. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/antopol> (дата обращения: 04.05.2013).
 14. Федотов А.М. Концептуальные подходы к построению распределенных систем // Труды Международной конференции по вычислительной математике (МКВМ-2004): Рабочие совещания. Новосибирск: ИВМиМГ СО РАН, 2004. - С.132-143.
 15. Шокин Ю.И., Федотов А.М. Поддержка и развитие распределенных информационно-вычислительных ресурсов в СО РАН // Вычислительные технологии. (Совместный выпуск). Вестн. КазНУ им. аль-Фараби. Серия: Математика, механика, информатика. Ч. 4. - 2004. - Т.42. - Ч.4. - № 3. - С.324-334.
 16. Земсков А.И., Шрайберг Я.Л. Электронные библиотеки: учеб. пособие для студентов ун-тов и вузов культуры и искусств и др. учеб. заведений. 3-е изд. М.: ГПНТБ России. 2004.
 17. Воройский Ф.С. Электронные и традиционные библиотеки – суть не одно и то же [Электронный ресурс]. Электронные библиотеки: рос. науч. электронный журн. Т.6. Вып.5. 2003. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part5/voroisky> (дата

обращения: 04.05.2010).

18. Акимов С.И., Елизаров А.М., Ершова Т.В., Когаловский М.Р., Федоров А.О., Хохлов Ю.Е. Научно-методическая поддержка разработки научных электронных библиотек [Электронный ресурс]. Электронные библиотеки: рос. науч. электронный журн. 2005. Т. 8. № 1. URL: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part1/AEEKFH>.
19. Вигурский К. В. Что такое электронная библиотека? Доклад на конференции "Информационные технологии в образовании - 2005". [Электронный ресурс]. <http://rd.feb-web.ru/library.htm>.
20. Федотов А.М., Барахнин В.Б., Жижимов О.Л., Федотова О.А. Технология создания корпоративных информационных систем учета трудов научных работников // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2011. Т. 9. № 2. С. 31-41.
21. Жижимов О.Л., Федотов А.М., Федотова О.А. Построение типовой модели информационной системы для работы с документами по научному наследию // Вестник Новосибирского государственного университета. Серия: Информационные технологии. 2012. Т. 10. № 2. С. 5-14.
22. Резниченко В.А, Проскудина Г.Ю., Кудим К.А. Концептуальная модель электронной библиотеки [Электронный текст] / В. А. Резниченко, Г. Ю. Проскудина, К. А. Кудим// Труды XI Всероссийской научной конференции RCDL'2009», Россия, г. Петрозаводск (Карелия), 17-21 сентября 2009 г.- С. 23-31
23. ISO-14721 Reference Model for an Open Archival Information System (OAIS), Draft Recommended Standard, CCSDS 650.0-P-1.1 (Pink Book) Issue 1.1 August 2009.
24. Candela L., Castelli D., Dobрева M., Ferro N., Ioanni-dis Y., Katifori H., Koutrika G., Meghini C., Pagano P., Ross S., Agosti M., Schuldt H., Soergel D. The DELOS Digital Library Reference Model Foundations for Digital Libraries. IST-2002-2.3.1.12. Technology-enhanced Learning and Access to Cultural Heritage. Version 0.98, December 2007.
25. Шокин Ю.И., Федотов А.М., Гуськов А.Е., Жижимов О.Л., Столяров С.В. Электронные библиотеки - путь интеграции информационных ресурсов Сибирского отделения РАН // Вестник КазНУ. Серия: Математика, механика, информатика. - 2005. - Алматы: КазНУ. - № 2. - С.115-127. - ISSN 1563-0285.
26. Федотов А.М. Методологии построения распределенных систем // Избранные доклады X Российской конференции «Распределенные информационно-вычислительные ресурсы» (DICR-2005), Новосибирск 6-8 октября 2005 г. / Вычислительные технологии. 2006. - Т.11. - С.3-

16. - Новосибирск: ИВТ СО РАН. - ISSN 1560-7534.
27. Otlet P. Traite de documentation. // Bruxelles: Ed. Mundaneum, 1934.
28. Отле П. Библиотека, библиография, документация: Избранные труды пионера информатики / Поль Отле. - Москва: ФАИР-ПРЕСС: Пашков Дом, 2004. - 348, [1] с. - (Специальный издательский проект для библиотек). - Библиогр.: с. 312-327. - Имен. указ.: с. 340-342. - ISBN 5-8183-0624-0
29. ГОСТ Р ИСО / МЭК ТО 10000-2-99. Информационная технология. Основы и таксономия функциональных стандартов. Часть 2. Принципы и таксономия профилей ВОС.
30. Бездушный А.Н., Бездушный А.А., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН. М.: Вычислительный Центр им. А.А. Дородницына РАН, 2006.
31. Федотов А.М., Барахнин В.Б., Жижимов О.Л., Федотова О.А. Модель информационной системы для поддержки научно-педагогической деятельности // Вестник НГУ. Сер.: Информационные технологии. - 2014. - Т.12. - № 1. - С.89-101. - ISSN 1818-7900.
32. Федотов А.М., Жижимов О.Л., Князева А.А., Колобов О.С., Мазов Н.А., Турчановский И.Ю., Федотова О.А. Проблемы авторитетного контроля для распределенных электронных библиотек и библиографических баз данных. Вестник НГУ. Серия: Информационные технологии. 2011. Т. 9. № 1. С.89–101.
33. Захаров А.А., Серебряков В.А. Система управления электронными библиотеками LibMeta // Труды 12-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – RCDL'2010, Казань, Россия, 2010. – с. 28-37.
34. DCMI – Dublin Core Metadata Initiative (<http://www.dublincore.org/>).
35. DSpace [Электронный ресурс] : an open source solution for accessing, managing and preserving scholarly works // dspace.org [web-сайт] / MIT Libraries; HP Labs. – 2007. <<http://www.dspace.org/>>
36. EPrints Free Software [Электронный ресурс] // EPrints for Digital Repositories [web-сайт] / School of Electronics and Computer Science, University of Southampton, UK. – 2008. <<http://www.eprints.org/>>
37. Fedora [Электронный ресурс] : Fedora Repository System // Fedora Commons [web-сайт] / Gordon and Betty Moor Foundation; Cornell University Information Science; University of Virginia Library; The Andrew W. Mellon Foundation. – 2007. <<http://www.fedora-commons.org/>>
38. Кудим К.А., Проскудина Г.Ю., Резниченко В.А. Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1 / Девятая всероссийская научная конференция «Электронные

- библиотеки: перспективные методы и технологии, электронные коллекции». Переяславль-Залесский, 15–18 октября 2007 года.
39. The Open Archives Initiative Protocol for Metadata Harvesting [Электронный ресурс]: Protocol Version 2.0 of 2002-06-14 // Open Archives Initiative: [web-сайт] / The OAI Executive; OAI Technical Committee. 2004 (<http://www.openarchives.org/>).
 40. RFC 4510 [Электронный ресурс]: Lightweight Directory Access Protocol (LDAP): Technical Specification Road Map / OpenLDAP Foundation. – 2006. <<http://www.apps.ietf.org/rfc/rfc4510.html>>
 41. ANSI/NISO Z39.50-2003. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. NISO Press, Bethesda, Maryland, U.S.A. November 2002.
 42. Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. ОИГТМ СО РАН, Новосибирск: ИВТ СО РАН. 2004. – 361 с.
 43. SRU (Search/Retrieve via URL) [Электронный ресурс]. Режим доступа: <http://www.loc.gov/standards/sru/> (дата обращения: 23.08.2013).
 44. Жижимов О.Л., Федотов А.М., Шокин Ю.И. Технологическая платформа массовой интеграции гетерогенных данных // Вестник НГУ. Серия: Информационные технологии. 2013. Т. 11. № 1. С. 24-41.
 45. ГОСТ 7.19-2001. Система стандартов по информации, библиотечному и издательскому делу. Формат для обмена данными. Содержание записи
 46. RUSMARC [Электронный ресурс]: Российский коммуникативный формат // Российская Библиотечная Ассоциация [web-сайт] / Российская Библиотечная Ассоциация. <http://www.rba.ru:8101/rusmarc/>
 47. Extensible Markup Language (XML) 1.0 (Fourth edition) [Электронный ресурс] // World Wide Web Consortium [web-сайт]. – 2006. <<http://www.w3.org/XML/>>
 48. ISO 2709:2008: Information and documentation – Format for information exchange // ISO - International Organization for Standardization. – 2008.