

Технологии создания распределенных информационных систем для поддержки научных исследований

Ю. И. Шокин, А. М. Федотов*, О. Л. Жижимов

Институт вычислительных технологий СО РАН, Новосибирск, Россия

*Контактный e-mail: fedotov@sbras.ru

Статья посвящена истории разработки и описанию технологических подходов, применяемых при создании распределенных информационных систем. Описываются разработанные архитектурные решения для информационных систем, предназначенных для поддержки научных исследований, и принципы интеграции их с внешними источниками. Определены функциональные требования к модели такой информационной системы, принципы ее организации, которые обуславливаются, во-первых, информационными потребностями исследователей, а во-вторых, необходимостью обеспечения надежного и долговременного хранения информации. Приведены правила представления и преобразования метаданных, а также описана работа со словарями, которые используются для систематизации и классификации информационных ресурсов и моделирования связей между ними.

Ключевые слова: информационная система, электронная библиотека, словарь-справочник, распределенные информационные ресурсы, классификация информационных ресурсов, интеграция данных, цифровой репозиторий, поиск информации, информационно-поисковый тезаурус, ключевой термин, метаданные, библиографическая база данных.

Введение

Одним из основных результатов созидательной, социальной и интеллектуальной человеческой деятельности является создание и накопление информационных ресурсов с целью их дальнейшего использования и сохранения опыта предыдущих поколений. Не будет преувеличением сказать, что уровень развития технологий сохранения информации и эффективности использования накопленной ранее информации на протяжении всей истории человечества значительно влиял на развитие производительных сил. Утеря информации приводила к отбрасыванию цивилизации на века назад. Однако, чтобы эффективно пользоваться накопленной ранее информацией, необходимы специальные инструменты и технологии, при помощи которых могут быть реализованы специальные приемы работы с информацией [1].

Стремительное развитие глобальных информационных и вычислительных сетей ведет к изменению фундаментальных парадигм обработки данных, которые можно охарактеризовать как переход к поддержке и развитию распределенных информационных

ресурсов [2–4]. Поэтому важнейшей задачей, связанной с технологией работы с информацией, является исследование способов интеграции распределенных источников данных и создание научного задела в области распределенных информационных систем и баз данных в целях разработки технологии, поддерживающей создание и функционирование широкомасштабных информационных инфраструктур на основе виртуальной интеграции. Такая технология позволит создавать глобальные инфраструктуры из десятков и сотен гетерогенных баз данных и решать стратегические задачи в области автоматизации различных форм распределенной деятельности. Более узкой целью является разработка принципов и программных средств виртуальной интеграции распределенных источников данных на основе международных стандартов и рекомендаций для создания масштабных информационных инфраструктур, предназначенных для виртуализации доступа к данным различных СУБД с использованием единых правил и политик [5].

На самом деле идея создания универсальной системы доступа к информационным ресурсам, распределенным в мировом пространстве, далеко не нова. По всей видимости, впервые ее четко осознал известный бельгийский ученый Поль Отле¹ в конце XIX в., предложив совершенно новый метод, названным им “Документацией”:

“Цели Документации состоят в том, чтобы суметь предложить документированные ответы на запросы по любому предмету в любой области знания: 1) универсальные по содержанию; 2) точные и истинные; 3) полные; 4) оперативные; 5) отражающие последние данные; 6) доступные; 7) заранее собранные и готовые к передаче; 8) предоставленные как можно большему числу людей” [6, с. 190].

“... человеческое знание позволит создать оборудование, действующее на расстоянии, в котором соединятся радио, рентгеновские лучи, кинематограф и микроскопическая фотография. Все предметы Вселенной, все предметы, созданные Человеком, будут регистрироваться на расстоянии с момента их создания. Тем самым будет создан движущийся образ мира — его память, его подлинная копия. Любой человек сможет прочесть отрывок, спроецированный на его личный экран” [6, с. 16].

Под интеграцией информационных ресурсов понимают их объединение с целью использования (с помощью удобных и унифицированных пользовательских интерфейсов) разнородной информации с сохранением ее свойств, особенностей представления и 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 позволяют работать с гетерогенными наборами и базами данных или системами баз данных, обеспечивая эффективность информационных поисков независимо от особенностей конкретных систем хранения ресурсов, к которым осуществляется доступ [5].

Исходя из анализа литературных источников и многолетней практики авторов в области создания программных комплексов для организации доступа к гетерогенным информационным ресурсам и базам данных наиболее оптимальной для платформы массовой интеграции баз данных представляется архитектура слабо связанных самодостаточных узлов распределенной информационной системы [4, 7–10]. Так, в ИВТ СО РАН разработана платформа ZooSPACE. Этимология этого названия основана на двух элементах. Элемент “SPACE” указывает на распределенность системы, которая

¹Поль Отле (1868–1944) — Paul Marie Ghislain Otlet — “отец” современной информатики.

создает некое пространство, в котором могут функционировать информационные узлы и сервисы, обеспечивая самосогласованный доступ к информационным ресурсам и базам данных, элемент “Zoo” — на некоторую преемственность предлагаемых решений по отношению к разработанным ранее программным комплексам в области обеспечения унифицированного доступа к гетерогенным базам данных. В первую очередь имеется в виду программный комплекс ZooPARK, разные версии которого успешно эксплуатируются в России и в ближнем зарубежье на протяжении последних 16 лет [11].

1. История вопроса

Разработка систем, интегрирующих информационные ресурсы, в ИВТ СО РАН началась около 20 лет назад еще в 1995 г. с создания информационных порталов института и Сибирского отделения РАН. Основные направления работ были связаны с формированием собственных электронных ресурсов по основным отраслям наук (математике, наукам о земле, химии, биологии, археологии и др.), созданием и поддержкой электронных коллекций и электронных публикаций, организацией удобных систем доступа к библиотечным и библиографическим базам данных и базам данных институтов Отделения [12, 13]. Результаты работ обеспечивали:

- единую распределенную информационную среду;
- информационную поддержку исследований по фундаментальным и прикладным направлениям;
- поддержку профессионально-ориентированных систем подготовки и обмена научными документами;
- поддержку профессионально-ориентированных систем доступа и интерфейсов с хранилищами данных;
- коллективное использование приобретаемой электронной литературы, каталогов, баз данных и библиографических изданий;
- поддержку электронных версий научных журналов, издаваемых институтами;
- поддержку принятия и реализации организационных и управленческих решений.

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 распределенной информационной системы (ИРИС) [2, 12, 13]. ИРИС СО РАН представляет собой полнофункциональную распределенную информационную систему, содержащую данные об интеллектуальном потенциале Отделения (институтах, сотрудниках, достижениях и пр.) и систему электронной поддержки сбора и накопления научной информации (электронных атласов, коллекций, баз данных и т. п.).

В основу этой информационной системы был положен WWW-сервер², который стал интегрирующим звеном для системы поддержки информационных ресурсов Отделения (рис. 1). Сервер содержал разнообразную информацию о деятельности Сибирского отделения, включая: базу данных организаций и сотрудников СО РАН (рис. 2); доску

²Сегодня он называется Порталом СО РАН (<http://www.sbras.ru/>). По данным Webometrics, (на 03.02.2015) сайт занимает первое место в России (<http://research.webometrics.info/en/Europe/Russian%20Federation>) среди научных организаций и 45-е место в мире (<http://research.webometrics.info/en/world>).

объявлений; основные результаты исследований и годовые отчеты СО РАН; систему поддержки конференций; электронную библиотеку СО РАН; важнейшие разработки институтов СО РАН, предлагаемые для широкого использования; информацию о сети Новосибирского научного центра; информационные проекты СО РАН; геоинформационные технологии в СО РАН; сведения о Доме ученых СО РАН; материалы газеты “Наука в Сибири” (рис. 3); справочные материалы по информационным ресурсам; нормативные документы РФ; фотоальбом и историю Сибирского отделения РАН и много других ресурсов.

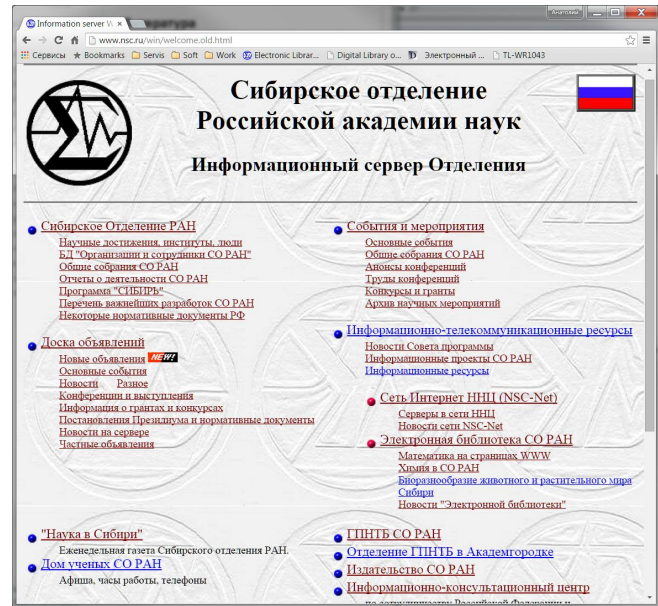
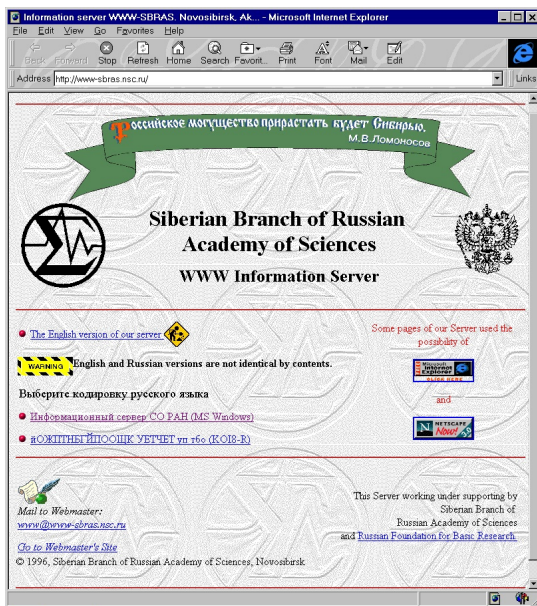


Рис. 1. Информационный сервер СО РАН

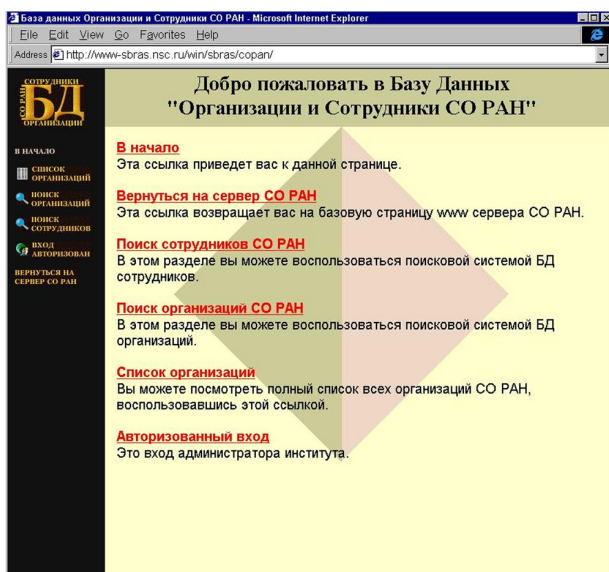


Рис. 2. База данных СО РАН

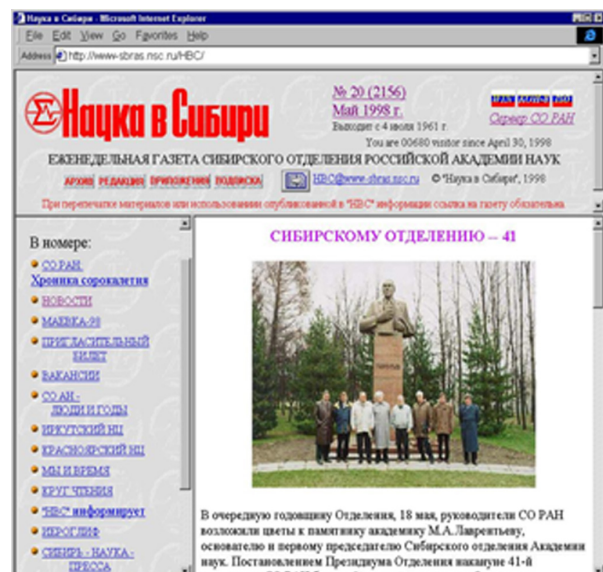


Рис. 3. Газета “Наука в Сибири”

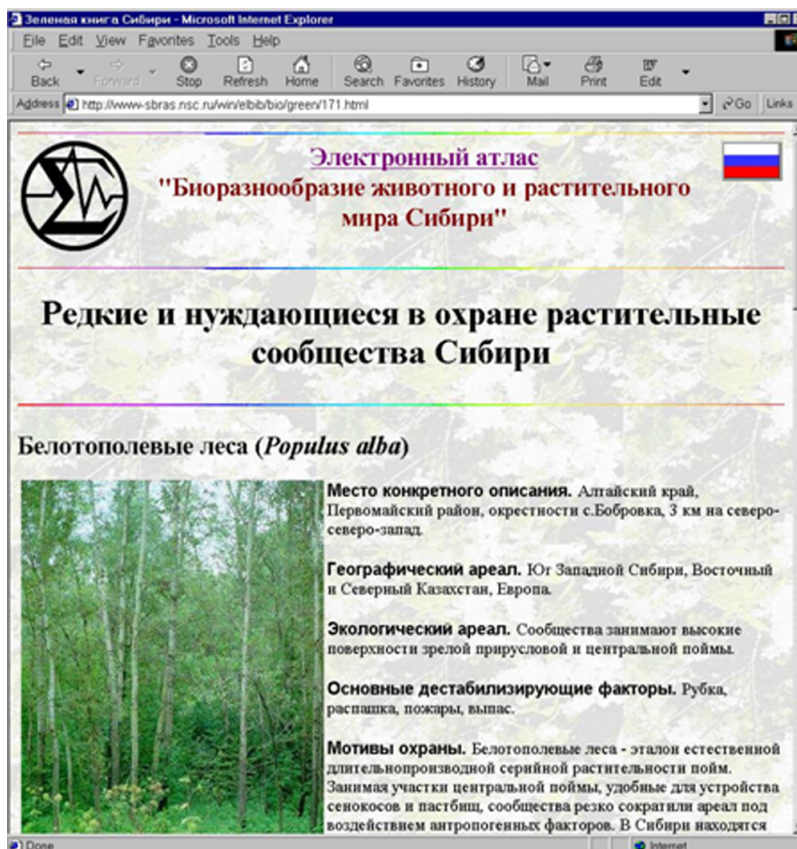


Рис. 4. Страница из электронного атласа “Биоразнообразие животного и растительного мира Сибири”

Некоторые созданные ресурсы не потеряли своей актуальности до сих пор: электронный атлас “Биоразнообразие животного и растительного мира Сибири” (рис. 4) [14], система поддержки проведения конференций³ [15], виртуальный музей СО РАН [16], база данных организаций и сотрудников СО РАН [17].

В результате работ по программе сложилось понимание того, что информационная система для поддержки научных исследований ИРИС должна основываться на концепции электронных (цифровых) библиотек [18–21]. В рамках такого подхода цифровые библиотеки рассматриваются как отдельная конкретная технология работы с цифровой информацией, образующая новый класс информационных систем (ИС), предназначенных для управления информационными ресурсами [22, 23]. Под термином “электронная библиотека” (ЭБ) здесь будем понимать систему управления структурированными каталогизированными коллекциями разнородных электронных (цифровых) объектов (ресурсов). Система управления информационными ресурсами не только обеспечивает поиск, навигацию по рубрикам (по словарям-классификаторам), но и предоставляет непосредственно пользователю конкретный ресурс (публикацию, документ, фотографию, описание факта и др.), а также дополнительные сведения о нем, например, географическую привязку, информацию об авторах, информацию о событиях, библиографию, перечень организаций и т. д.

³Отметим, что первая конференция при поддержке электронной системы проведена ИВТ СО РАН в 1996 г. по электронным публикациям EL-PUB-96 (<http://www.ict.nsc.ru/ws/elpub96/>).

2. Концепция ИРИС

Наиболее эффективным способом решения проблем организации доступа к распределенным информационным ресурсам является организация информации о них в информационные системы в виде электронных библиотек. Иными словами, понятие электронной библиотеки подразумевает, что любой ресурс должен быть стандартным образом описан (каталогизирован), снабжен метаданными, правилами доступа и уникальным идентификатором. В работах [24–26] были сформулированы основные принципы реализации ИРИС, основанные на использовании идеи электронных библиотек.

Отметим, что ЭБ — явление относительно новое, но уже достаточно популярное [27]. Тем не менее ЭБ сегодня следует рассматривать как множество слабо связанных сущностей, объединяемых на первый взгляд только общим названием [23, 28]. Современное название “электронная библиотека” — это не только и не столько дань моде, сколько попытка охарактеризовать новый феномен — принципиально новый класс систем, призванных аккумулировать и распространять информацию в электронной форме [29]. Большой интерес к системам данного класса объясняется потребностями общества и ростом возможностей по их удовлетворению [30].

Сформулируем основные цели, стоящие перед ЭБ (системами управления информационными ресурсами) [31]:

- 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 непригодно для пользователя-системы. Для реализации функций поиска вне графических интерфейсов требуется наличие специальных сетевых сервисов и языков запросов. В идеальном случае все ИС должны поддерживать единый поисковый профиль и единый язык запросов [22].

К “электронным библиотекам” могут относиться различные объекты, такие как архивы цифрового контента и наборы программного обеспечения для управления этим контентом. Электронной библиотекой может называться система сетевых сервисов, предоставляющих доступ к цифровому контенту и объединенных единой системой управления этим доступом [31]. Такое определение ЭБ полностью соответствует определению традиционной библиотеки как организации в системе [22].

Ввиду того, что информация в ИС отображает некоторые сущности реального мира (предметы, процессы, явления, персоны, публикации, документы, алгоритмы, программы, файлы, факты, ключевые термины и т. д.), информационные системы следует рассматривать как множество информационных объектов (наборов данных), 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 [32]. Обзор существующих концептуальных моделей ЭБ приведен в [33].

Онтологическая модель ЭБ ИРИС основана на концептуальных моделях электронных библиотек RM OAIS [34] и DELOS DLRM [35].

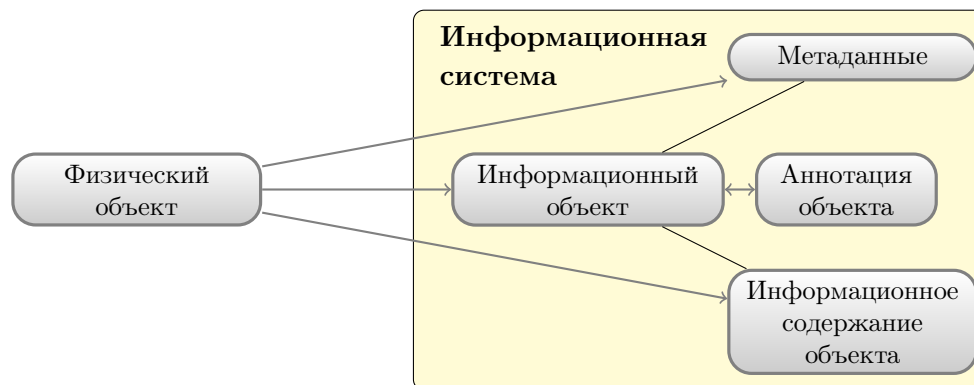


Рис. 5. Структура сущностей информационной системы

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 характеризуется такими особенностями:

- имеет идентификатор;
- 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;
- может регулироваться функциями, управляющими его жизненным циклом;
- выражается через информационный объект;
- должен быть описан метаданными, а также аннотациями.

В основе реализации системы управления электронными библиотеками (СУЭБ) в ИРИС лежит метамодель, исходящая из того, что каждый информационный ресурс имеет набор присущих ему атрибутов и методов, характеризующих его свойства и связи с другими ресурсами. Эффективное средство описания информационных объектов — метаданные, которые являются неотъемлемой частью информационного объекта и описывают реальный объект или группу объектов.

Каждый информационный объект в ИС включает (рис. 5):

- информационное содержание объекта — первичный информационный объект (например, изображение, полный текст и т. д.), который может использоваться самостоятельно;
- метаданные — объект, содержащий описание первичного информационного объекта в стандартизированной форме;
- аннотацию — объект, описывающий использование ИР или его части. Аннотации могут включать примечания, структурированные комментарии и связи. Объекты аннотации помогают интерпретировать ИР, содержат либо поддержку, либо детальные объяснения, либо информацию о том, как можно использовать ИР.

3. Документы

В информационном пространстве все ресурсы (события, факты, программы и любые другие сущности реального или виртуального мира) существуют только в форме некоторых информационных объектов [4, 36].

Бельгийский ученый Поль Отле — пионер и основоположник науки “информатика” — в своем трактате о документации [37] определяет понятие “документ” как материальный объект, содержащий информацию, специально предназначенный для ее передачи в пространстве и во времени, который трактуется как основной “объект” и с которым оперирует любая информационная система [6]. Таким образом, документ — это информационный объект, представляющий собой структурированное описание реальной сущности (объекта, субъекта, факта или понятия). Совокупность объектов составляет 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. Документом могут быть статья из журнала, сам журнал, данные о персоне, оцифрованное изображение, экспериментальные данные, программа или вычислительный алгоритм, база данных, фрагмент базы данных и т. п. [4]. Например, документ может не содержать полный текст статьи, но в качестве атрибута может иметь указатель на хранилище данных (репозиторий), где хранится полный текст. Документ должен иметь уникальный идентификатор, структурированное описание (метаданные), набор атрибутов (свойств) и методов (функций). Взаимодействие одного документа с другим (например, работа с атрибутами) происходит через набор методов (функций).

Ключевым моментом в работе с документами является использование метаданных для формирования структуры, схемы данных и свойств документов — информационных элементов системы и ведения каталога системы. Система управления ИС опирается на метаинформацию, содержащую онтологию, определяющую принципы организации информации. Онтология конкретной предметной области обуславливает схему данных (атрибутов) для метаданных [4].

4. Электронные библиотеки

Электронные библиотеки — это каталогизированные распределенные информационные системы, позволяющие хранить, обрабатывать, распространять, анализировать данные, а также организовывать их поиск в разнообразных коллекциях электронных (цифровых) документов. Основные задачи, решаемые электронными библиотеками, — это управление информационными ресурсами и интеграция информационных ресурсов (включая поддержку унифицированного доступа к ним), а также эффективная навигация в них [36].

Под интеграцией информационных ресурсов понимают их объединение с сохранением свойств, особенностей представления и возможностей манипулирования с ними. 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 позволяют работать с гетерогенными наборами и базами данных или системами баз данных, обеспечивая пользователю эффективность информационного поиска независимо от особенностей конкретных систем хранения ресурсов, к которым осуществляется доступ [4].

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

странстве. При таком подходе хорошо известные информационно-поисковые системы, используемые в информационных системах и базах данных, являются частными случаями навигационных средств [4, 36].

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 понимают 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 [38]. Наиболее важным является профиль метаданных информации, циркулирующей в системе. Выбор профиля метаданных должен основываться на выполнении следующих требований [4, 31, 39, 40]:

- наличие описаний основных типов информации, необходимой для поддержки научно-образовательной деятельности;
- открытость, т. е. он должен обеспечивать доступ к информации в соответствии с ее описанием (метаданными);
- расширяемость, т. е. возможность детализации описаний;
- возможность интеграции информации;
- возможность уникальной идентификации информации;
- обеспечение отбора, систематизации и классификации информации;
- возможность размещения и поиска информации в распределенной среде;
- соответствие современным технологиям описания и использования информации;
- интероперабельность с другими системами.

Серьезной проблемой является идентификация информационных ресурсов [41], позволяющая получать библиографические сведения, а также устанавливать связи определенного ресурса с другими фактами и объектами. При работе с цифровыми объектами принят определенный набор стереотипов [22], например взаимные ссылки между цифровыми объектами (в виде гиперсвязей в пользовательских графических интерфейсах просмотра информации). Реализация взаимных ссылок в цифровых документах не представляет большой сложности, однако имеет свою специфику. 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 может быть замена жестких гиперссылок динамическими ассоциативными связями между документами (или между элементами, составляющими документ), определяемыми в момент представления документа пользователю, на уровне системы управления.

Существует достаточно много технологических разработок ИС для электронных библиотек, так или иначе ориентированных на поддержку научных исследований, например, euroCRIS (<http://www.eurocris.org/>), eLibrary (<http://elibrary.ru/>), Информика (<http://www.informika.ru/>), MathNET (<http://www.mathnet.ru/>). Информационным потребностям научно-образовательного сообщества в информации в большей степени удовлетворяет система ИСИР (ЕНИИ) РАН [39, 42].

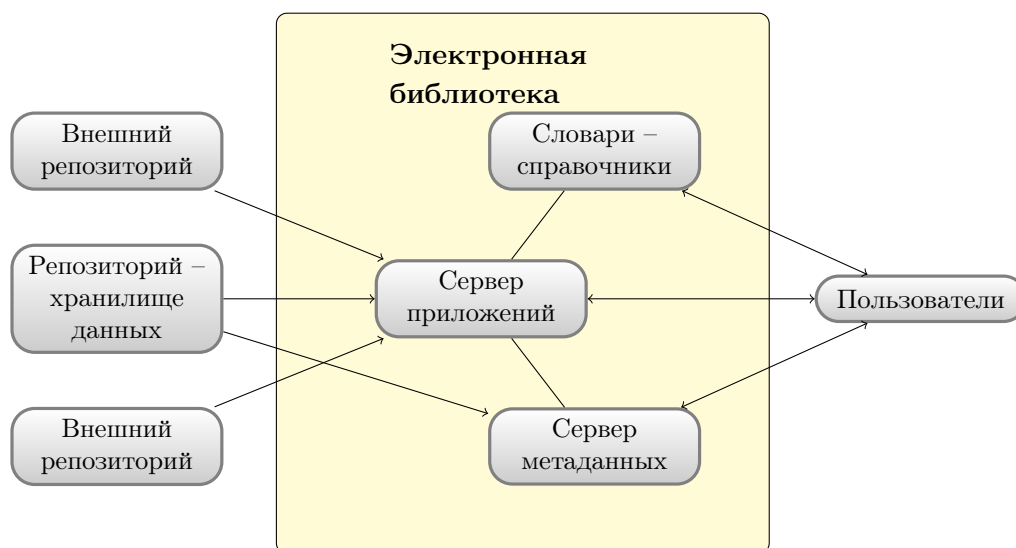


Рис. 6. Архитектура электронной библиотеки

Таким образом, информационная модель ЭБ должна быть многоуровневой и состоять как минимум из нескольких компонентов: хранилища данных (репозитория), сервера метаданных, сервера приложений (диспетчера), словарей-справочников (рис. 6) [4, 32, 36].

5. Цифровые репозитории

Хранилище данных (цифровой репозиторий) является одним из важнейших компонентов распределенной системы и предназначен только для обеспечения “функции” долговременного хранения информационных ресурсов. Может измениться система, могут поменяться интерфейсы и сервисы, но ресурс (документ), несущий информацию, не изменяется, поэтому должен храниться вечно и независимо, т.е. функция хранения данных не зависит от других функций и сервисов системы.

Для организации системы долговременного хранения информационных ресурсов (репозитория цифровых объектов) международной организацией по стандартизации (ISO) предложен стандарт ISO-14721 (Open Archive Information System — OAIS) [34]. Эталонная модель для стандарта OAIS — это концептуальная модель, основанная на расширенной схеме данных Dublin Core [43]. Эта модель использована многими организациями для разработки наборов метаданных и создания крупных хранилищ цифровых объектов. На ее основе 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 особенностям институционального репозитория относятся:

- обеспечение разграниченного доступа к разнородным цифровым объектам (публикациям, изображениям и т. д.);
- организация доступа к информационным ресурсам со стороны 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 OpenDOAR⁴, большинство институциональных репозиторий основано на свободном программном обеспечении и построено в рамках модели OAIS на базе технологий открытых систем. В мире насчитывается более десятка систем поддержки институциональных репозиторий, наиболее популярные из них DSpace [44] (свыше 41 % установок), E-Prints [45], Fedora [46]. Сравнительную характеристику этих систем и описание используемых в них информационных моделей можно найти в [47]. 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) [48], который поддерживается большинством систем, предназначенных для хранения информационных ресурсов.

В качестве основного репозитория в ИВТ СО РАН была выбрана система DSpace как самая популярная в мире и уже на протяжении десяти лет эксплуатирующаяся в СО РАН (а также в ряде других институтов и университетов России). Система хранит информацию о пользователях, поддерживает авторизацию и разграничивает доступ к содержимому репозитория по группам, сетевым адресам и на основе протокола LDAP [49], что при создании ИС дает возможность использовать уже существующую систему аутентификации пользователей (а не разрабатывать свою собственную) и достаточно легко дифференцировать публичные и служебные ресурсы, оставляя при этом свободный доступ к метаданным. Наличие провайдеров данных для протоколов OAI-PMH [48], Z39.50 [9, 50] и SRW/SRU [51] позволяет разрабатывать программный интерфейс для взаимодействия различных ИС с хранилищем данных, построенным на основе DSpace.

6. Выбор метаданных

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 понимают 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 информационными системами требуются специальные технологические методы и общие соглашения. Это влечет за собой необходимость соблюдения соответствия всех схем данных, интерфейсов и протоколов международным стандартам и рекомендациям [22, 31]. В работах [31, 32] был определен профиль ЭБ как необходимый набор стандартов и компонентов информационной системы, ориентированной на научные исследования.

⁴The Directory of Open Access Repositories — http://www.open_doar.org/

Интеграция распределенных информационных систем основана на метаданных, представляемых в соответствии с унифицированными данными (профилем информационной системы). Под интеграцией данных с точки зрения пользователя следует понимать возможность свободно группировать любые имеющиеся разнородные данные по любому признаку в произвольные реальные и/или виртуальные коллекции и организовывать по всем массивам данных прозрачный для конечного потребителя сквозной поиск информации.

Реализация механизмов интеграции данных немыслима без их стандартизации — данные одного типа должны описываться единым образом в соответствии с нормативными документами. Так, в стандартизованном виде должны предоставляться следующие типы информационных ресурсов:

- географические материалы (карты, спутниковые снимки, данные полевых наблюдений и т. п.), а также соответствующие базы метаданных;
- фактографические базы данных и метаданных;
- библиографические базы данных и электронные каталоги;
- полнотекстовые базы данных и цифровые репозитории;
- авторитетные базы данных (словари, справочники и т. п.);
- другие ресурсы (аудио- и видеозаписи, электронные презентации и др.), снабженные стандартизованными метаданными.

Для обеспечения своей функциональности интегрированная информационная система должна содержать следующие подсистемы:

- идентификации информационных ресурсов;
- идентификации, аутентификации и авторизации пользователей;
- управления метаданными;
- управления информационными ресурсами;
- сбора статистики;
- мониторинга доступности сервисов и ресурсов.

Реализация подсистем ИС должна основываться на открытых спецификациях, связанных с международными стандартами и рекомендациями. В распределенной среде должны быть задействованы механизмы синхронизации данных, например, на основе репликаций (рис. 7). При этом в качестве протоколов сетевого взаимодействия должны выступать стандартные протоколы, такие как: OAI-PMH, OAI-ORE [48], SRW/SRU [51], Z39.50 [50], LDAP [49] и др.

Метаданные необходимы для решения следующих задач:

- 1) предоставление сведений о документах, их содержании, структуре, способах использования и т. д.;
- 2) систематизация и классификация документов;
- 3) организация процедур внутрисистемной обработки;
- 4) поддержка обмена с внешними ИС.

Метаданные подразделяют на следующие классы (рис. 8).

Административные или служебные метаданные. Содержат исключительно служебную информацию, например дату модификации документа, фамилию владельца документа (не путать с автором), права доступа к документу и т. п.

Системные метаданные. Обеспечивают технологические задачи системы управления ресурсами, например, содержат правила предоставления документов пользователю, правила преобразования схем данных структурных метаданных, правила определения ассоциативных связей между документами и т. п.

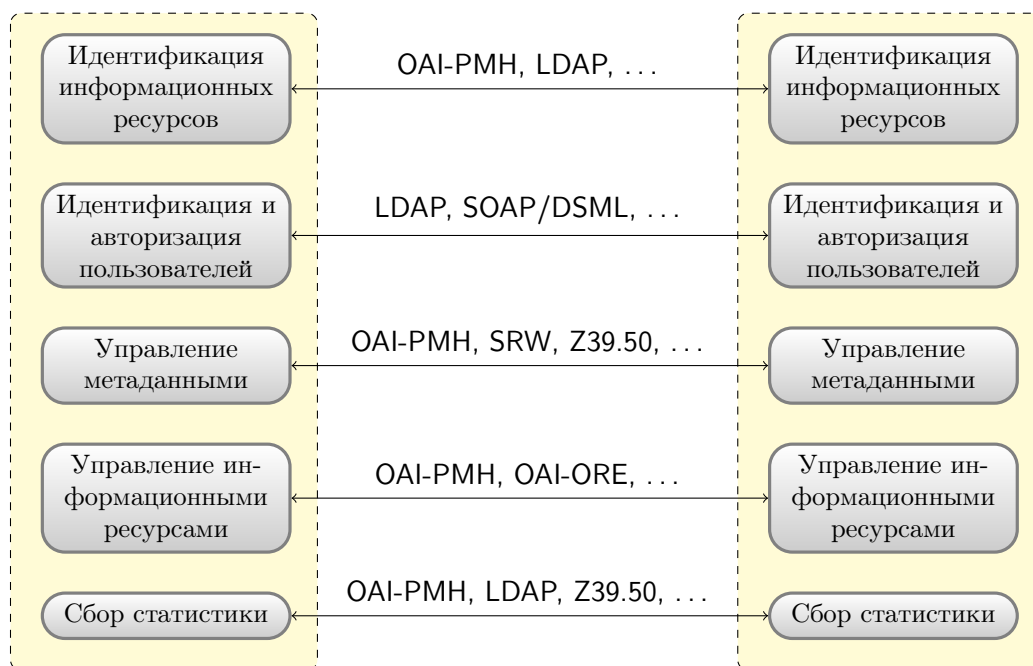


Рис. 7. Протоколы сетевого взаимодействия подсистем ИС

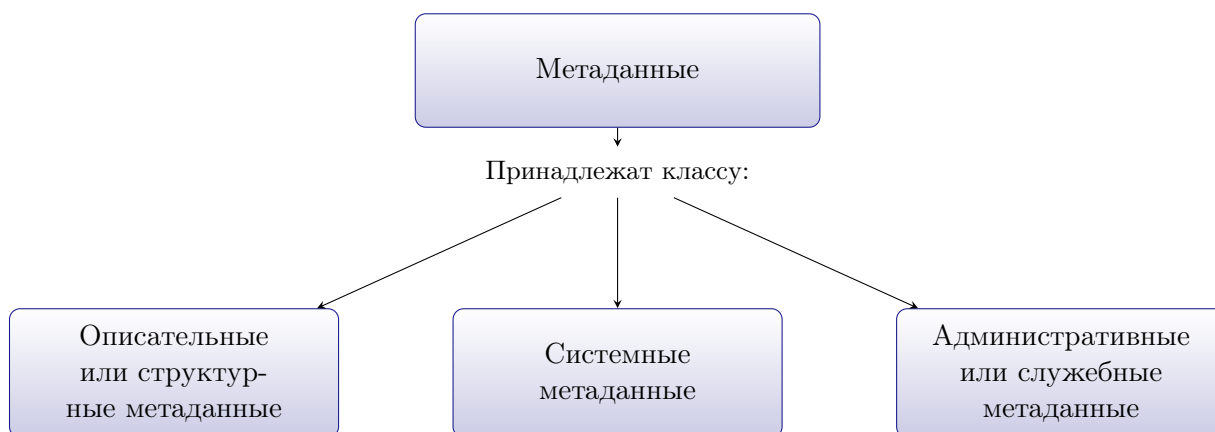


Рис. 8. Классы метаданных ИС

Описательные или структурные метаданные. Описывают документ в соответствии с выбранной схемой данных, например Dublin Core или МЕКОФ [52].

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

- субъекты (персоны, организации и т. д.);
- объекты или единицы хранения (публикации, документы, факты, научные результаты, мероприятия, фотографии и др.);
- отношения (понятия, ключевые термины, события, время, место и т. п.).

При этом указание на субъекты дается с помощью ссылки на экземпляр сущности “субъект”, что позволяет корректно решать задачу идентификации объектов.

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 производятся систематизация и классификация материала.

Для формирования метаданных применяется несколько стандартов, являющихся расширениями рекомендаций Dublin Core и Qualified Dublin Core (QDC). Для документов в ИВТ СО РАН стандартная схема метаданных была расширена QDC-полями, включающими основные требования государственного стандарта МЕКОФ [52].

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

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

7. Практическая реализация

Рассмотренная модель ИС реализована в виде системы управления электронными библиотеками (СУЭБ ИРИС), созданной и эксплуатируемой в ИВТ СО РАН с 2002 г. [36]. Конечно, за это время были проведены некоторые технологические модификации, но принципы, заложенные в начале разработки, сохранились и поддерживаются до сих пор. Главное достоинство данной технологии состоит в том, что она позволила сохранить все созданные ресурсы в актуальном виде и поддерживает к ним доступ. Может быть, это является главной причиной высокого рейтинга сайта СО РАН (рис. 9).

СУЭБ ИРИС оперирует электронными коллекциями документов. Электронная коллекция — это набор документов, объединенных по смысловому признаку и имеющих одинаковую структуру (схему данных). Вся работа системы строится на использовании ряда технологий (рис. 10). Пользователям и приложениям предоставляется полный набор услуг (пользовательских и административных) по доступу к информационным ресурсам системы (рис. 11).

СУЭБ позволяет работать с двумя видами коллекций: каталогами и тезаурусами. Принципиальное отличие каталогов от тезаурусов состоит в том, что в тезаурусах можно организовать иерархические зависимости (родитель — потомок, часть — целое и т. п.) между записями. Коллекции-каталоги предназначены для хранения и обработки метаданных о документах различной природы: публикациях, ключевых терминах, персонах, организациях, фотографиях и т. д. Коллекции-тезаурусы предназначены для работы со словарями-классификаторами.

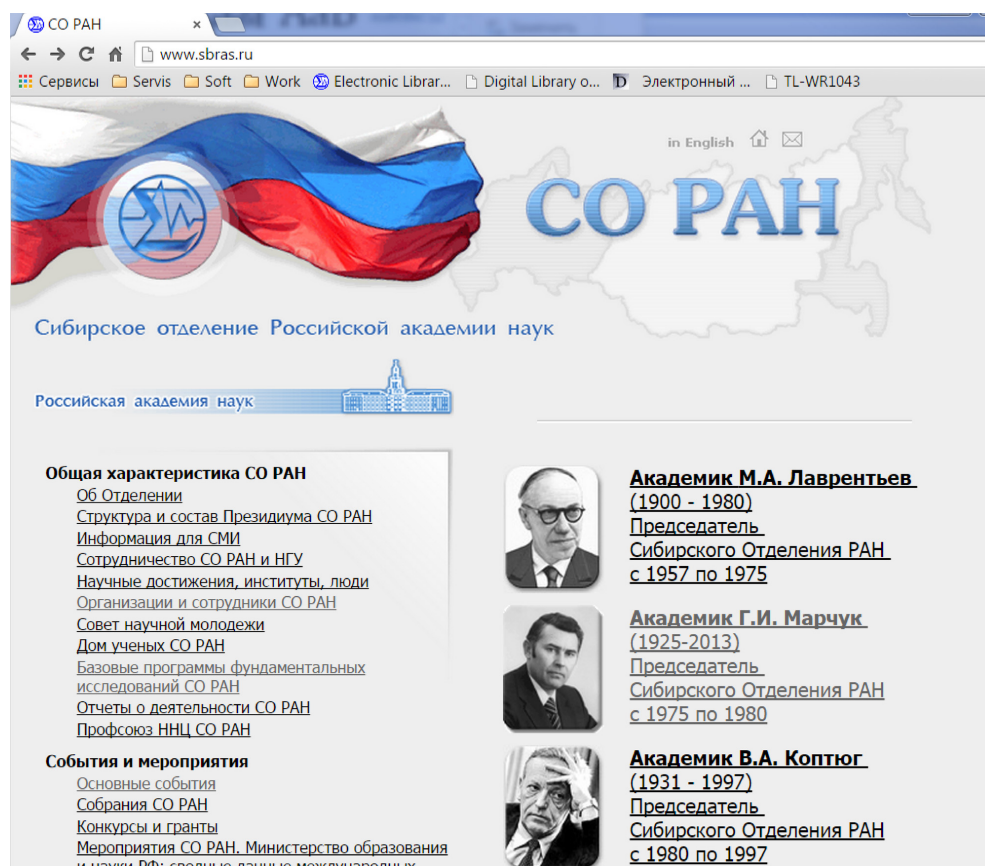


Рис. 9. Информационный сервер СО РАН 2004–2015 гг.

Сервер метаданных СУЭБ имеет служебную коллекцию “Основной каталог метаданных”, которая содержит документы, описывающие все метаданные, которые можно использовать в системе. Документы основного каталога содержат описания схемы QDC, расширенной элементами для соответствия МЕКОФ, и служебные метаданные, описывающие структуру объектов, пользовательские интерфейсы, ассоциативные связи между документами, права доступа к документам и т. д. При желании набор 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 основного каталога.

В СУЭБ представлены 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 двумя способами.

1. Ссылка на идентификатор записи — уникальный, в пределах одной коллекции, текстовый код, формируемый в рамках конкретной коллекции по определенным пра-

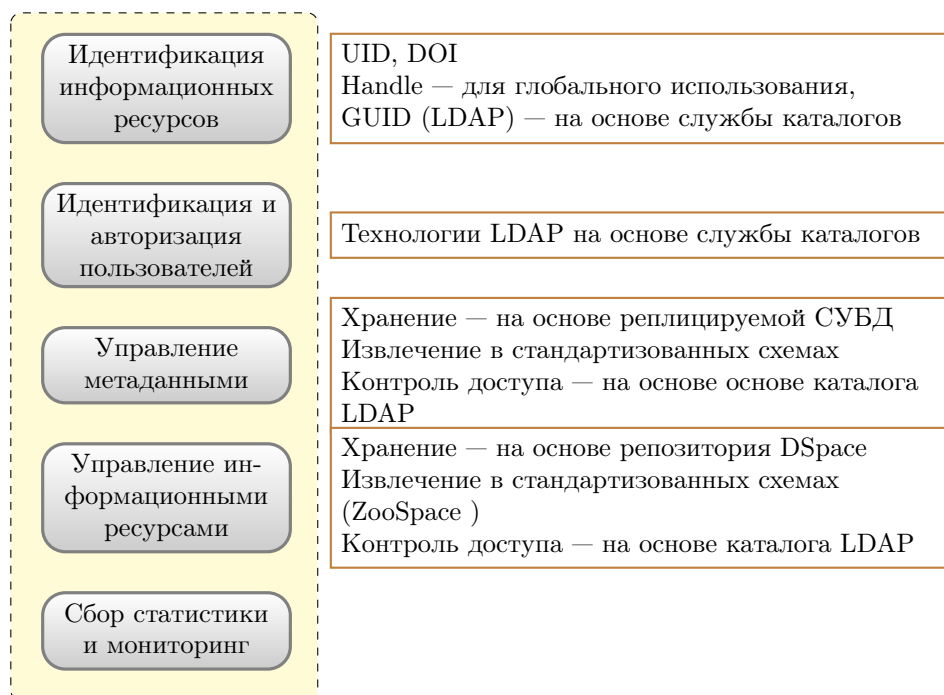


Рис. 10. Используемые технологии реализации ИС



Рис. 11. Структура интерфейсов ИС

вилам. 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.

2. 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 ключевой термин присутствует в метаданных.

С целью организации обмена метаданными между репозиториями и сервером метаданных (а также другими системами с расширенным профилем) создан специальный сервис, выполняющий преобразование метаданных из внутренней схемы в другие схемы метаданных, в том числе и в схему DСMІ, с использованием квалификаторов (QDC), а также в схему МЕКОФ (представление ISO-2709 или XML).

Реализован сервис OAI-PMH, 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 применяются контролируемые словари из справочного блока сопровождения. Для обеспечения интероперабельности данных также использован сервер приложений на основе ZooPARK-ZS [5], реализующий доступ к метаданным системы по протоколам Z39.50 [50] и SRW/SRU [51].

Разработанная модель может быть использована как типовая модель системы для работы с документами, связанными с научно-образовательной деятельностью, поскольку решает основные задачи, предъявляемые к таким системам: обеспечение 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 и Z39.50.

Рассмотренная технология создания и поддержки информационных ресурсов кроме работы с научными коллекциями с успехом была реализована в научно-образовательной сфере на примере ЭБ научной школы Алексея Андреевича Ляпунова — основателя теоретического программирования и российской кибернетики, а также в виде ЭБ учебных пособий по курсам “Современные проблемы информатики и вычислительной техники”, “Вычислительные системы”, “Информатика” и “Экология” и др.

Список литературы / References

- [1] **Шокин Ю.И., Федотов А.М., Баракнин В.Б.** Проблемы поиска информации. Новосибирск: Наука, 2010. 198 с.
Shokin, Yu.I., Fedotov, A.M., Barakhnin, V.B. Problems of information retrieval. Novosibirsk: Nauka, 2010. 198 p. (in Russ.)
- [2] **Шокин Ю.И., Федотов А.М.** Распределенные информационные системы // Вычисл. технологии. 1998. Т. 3, № 5. С. 79–93.
Shokin, Yu.I., Fedotov, A.M. Distributed informational systems // Computational Technologies. 1998. Vol. 3, No. 5. P. 79–93. (in Russ.)
- [3] **Жижимов О.Л., Федотов А.М., Чубаров Л.Б., Шокин Ю.И.** Технология создания распределенных информационно-вычислительных ресурсов СО РАН // Тр. Первой Международной конференции САИТ-2005. Переславль-Залесский, 12–16 сентября 2005 г. Т. 2. М., 2005. С. 161–165.
Zhizhimov, O.L., Fedotov, A.M., Chubarov, L.B., Shokin, Yu.I. Technology of creation of distributed information resources of SB RAS // Proceedings First International Conference SAIT-2005. Pereslavl'-Zalesskiy, September 12–16, 2005. Vol. 2. Moscow, 2005. P. 161–165. (in Russ.)
- [4] **Федотов А.М.** Методологии построения распределенных систем // Вычисл. технологии. 2006. Т. 11. Спецвыпуск: Избранные доклады X Российской конференции “Распределенные информационно-вычислительные ресурсы” (DICR-2005). С. 3–16.

- Fedotov, A.M.** Methodologies of the distributed systems // Computational Technologies. 2006. Vol. 11. Special issue: Proceedings of X Russian Conference “The distributed information-computational resources” (DICR-2005). P. 3–16. (in Russ.)
- [5] **Жижимов О.Л., Федотов А.М., Шокин Ю.И.** Технологическая платформа массовой интеграции гетерогенных данных // Вестн. НГУ. Информационные технологии. 2013. Т. 11, вып. 1. С. 24–41.
Zhizhimov, O.L., Fedotov, A.M., Shokin, Yu.I. Technology platform for the mass integration of heterogeneous data // Vestnik NGU. Informatsionnye tekhnologii. 2013. Vol. 11, iss. 1. P. 24–41. (in Russ.)
- [6] **Отле П.** Библиотека, библиография, документация: Избранные труды пионера информатики. М.: ФАИР-ПРЕСС: Пашков Дом, 2004. 348 с.
Otle, P. Library, bibliography, documentation: Selected works of the pioneer computer science. Moscow: FAIR-PRESS: Pashkov House, 2004. 348 p. (in Russ.)
- [7] **Жижимов О.Л., Пестунов И.А., Федотов А.М.** Структура сервисов управления метаданными для разнородных информационных систем // Электронные библиотеки. 2012. Т. 15, № 6. Адрес доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2012/part6/ZPF>
Zhizhimov, O.L., Pestunov, I.A., Fedotov, A.M. Structure of metadata services management for heterogeneous information systems // Russian Digital Libraries Journal. 2012. Vol. 15, No. 6. Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2012/part6/ZPF> (in Russ.)
- [8] **Жижимов О.Л., Амелъченко С.А.** Информационная система проекта “Электронная Сибирь”: сервисы управления данными // Вест. ДВО РАН. 2012. № 2. С. 123–128.
Zhizhimov, O.L., Amelchenko, S.A. Information System of the project “Electronic Siberia”: data control services // Bulletin of the Far Eastern Branch of the Russian Academy of Sciences. 2012. No. 2. P. 123–128. (in Russ.)
- [9] **Жижимов О.Л., Мазов Н.А.** Принципы построения распределенных информационных систем на основе протокола Z39.50. Новосибирск: ОИГГМ СО РАН, 2004. 361 с.
Zhizhimov, O.L., Mazov, N.A. Principles of distributed information systems based on the protocol Z39.50. Novosibirsk: OIGGM SO RAN, 2004. 361 p. (in Russ.)
- [10] **Шокин Ю.И., Федотов А.М., Жижимов О.Л.** Технология распределенных информационных систем // Современные информационные технологии для научных исследований: Матер. Всерос. конф., Магадан, 20–24 апреля 2008 г. Магадан: СВНЦ ДВО РАН, 2008. С. 18–21.
Shokin, Yu.I., Fedotov, A.M., Zhizhimov, O.L. Technology of distributed information systems // Materialy Vserossiyskoy konferentsii “Sovremennye informatsionnye tekhnologii dlya nauchnykh issledovaniy”, Magadan, 20–24 April 2008. Magadan: SVNTs DVO RAN, 2008. P. 18–21. (in Russ.)
- [11] **Жижимов О.Л., Мазов Н.А.** Серверный комплекс ZooPARK — итог 10-летней эксплуатации [Электронный ресурс] // XVI Междунар. конф. “Библиотеки и информационные ресурсы в современном мире науки, культуры, образования и бизнеса”. Крым-2009. Судак, Украина, 08.06.–12.06.2009: Материалы конференции. М.: ГПНТБ России, 2009. Адрес доступа: <http://www.gpntb.ru/win/inter-events/crimea2009/disk/118.pdf>
Zhizhimov, O.L., Mazov, N.A. Server complex ZooPARK a total of 10 years of operation [Electronic resource] // XVI Mezhdunarodnaya konferentsiya “Biblioteki i informatsionnye resursy v sovremennom mire nauki, kul'tury, obrazovaniya i biznesa”. Sudak, Ukraina, 08.06–12.06.2009: Materialy konferentsii. Moscow: GPNTB Rossii, 2009. Available at: <http://www.gpntb.ru/win/inter-events/crimea2009/disk/118.pdf> (in Russ.).

- [12] **Шокин Ю.И., Федотов А.М.** Электронная библиотека Сибирского отделения РАН // Электронные библиотеки. 1999. Т. 2, вып. 4. Адрес доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part4/fedotov> (дата обращения: 04.09.2015).
Shokin, Yu.I., Fedotov, A.M. Electronic Library of the Siberian Branch of the Russian Academy of Sciences // Russian Digital Libraries Journal. 1999. Vol. 2, iss. 4. Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part4/fedotov> (accessed: 04.09.2015). (in Russ.)
- [13] **Шокин Ю.И., Федотов А.М., Жижимов О.Л., Мазов Н.А.** Интегрированная распределенная информационная система (ИРИС) Сибирского отделения РАН // Материалы выездного заседания научно-координационного совета по целевой программе “Информационно-телекоммуникационные ресурсы СО РАН”, Иркутск, 29–30 августа 2002 г. Иркутск: Ин-т географии им. В.Б. Сочавы СО РАН, 2003. С. 139–149.
Shokin, Yu.I., Fedotov, A.M., Zhizhimov, O.L., Mazov, N.A. Distributed Integrated Information System (IRIS), Siberian Branch of Russian Academy of Sciences // Proceedings of the visiting session of the Coordinating Council target program “Information and Telecommunication Resources SB RAS”, Irkutsk, August 29–30, 2002. Irkutsk: V.B. Sochava Institute of Geography SB RAS, 2003. P. 139–149. (in Russ.)
- [14] **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” // Вычисл. технологии. 1998. Т. 3, № 5. С. 68–78.
Fedotov, A.M., Artyomov, I.A., Ermakov, N.B., Krasnikov, A.A., Potyomkin, O.N., Ryabko, B.Ya, Fedotov, A.A., Khorev, A.G. Electronic atlas “Biological variety of the Siberian flora” // Computational Technologies. 1998. Vol. 3, No. 5. P. 68–78. (in Russ.)
- [15] **Федотов А.М., Гуськов А.Е., Молородов Ю.И.** Информационная система поддержки проведения конференций СО РАН // Материалы выездного заседания научно-координационного совета по целевой программе “Информационно-телекоммуникационные ресурсы СО РАН”, Иркутск, 29–30 августа 2002 г. Иркутск: Ин-т географии им. В.Б. Сочавы СО РАН, 2003. С. 91–110.
Fedotov A.M., Guskov A.E., Molorodov Yu.I. Information Support System conferences SB RAS // Proceedings of the visiting session of the Coordinating Council target program “Information and Telecommunication Resources SB RAS”, Irkutsk, August 29–30, 2002. Irkutsk: V.B. Sochava Institute of Geography SB RAS, 2003. P. 91–110. (in Russ.)
- [16] **Шокин Ю.И., Ламин В.А., Федотов А.М., Баракнин В.Б., Жижимов О.Л., Мазов Н.А., Пищик Б.Н., Покровский Н.Н., Рычкова Е.В.** Виртуальный музей Науки и Техники СО РАН // Материалы выездного заседания научно-координационного совета по целевой программе “Информационно-телекоммуникационные ресурсы СО РАН”, Иркутск, 29–30 августа 2002 г. Иркутск: Ин-т географии им. В.Б. Сочавы СО РАН, 2003. С. 118–125. **Shokin, Yu.I., Lamin, V.A., Fedotov, A.M., Barakhnin, V.B., Zhizhimov, O.L., Mazov, N.A., Pishchik, V.N., Pokrovskiy, N.N., Rychkova, E.V.** Virtual museum of science and technique SB RAS // Proceedings of the visiting session of the Coordinating Council target program “Information and Telecommunication Resources SB RAS”, Irkutsk, August 29–30, 2002. Irkutsk: V.B. Sochava Institute of Geography SB RAS, 2003. P. 118–125. (in Russ.)
- [17] **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.
Leonova, Yu.V., Klimenko, O.A., Fedotov, A.M. Information system “Database organizations and employees of SB RAS”. Novosibirsk: “RITS Prays-Kur’erp”, 2005. 55 p. (in Russ.)

- [18] **Федотов А.М., Шокин Ю.И.** Электронная библиотека Сибирского отделения РАН // Информационное общество. 2000. № 2. С. 22–31.
Fedotov, A.M., Shokin, Yu.I. Electronic library of the Siberian Branch of the Russian Academy of Sciences // Information Society. 2000. No 2. P. 22–31. (in Russ.)
- [19] **Шокин Ю.И., Федотов А.М.** Библиотека, работающая круглосуточно // ЭКО. 2000. № 6. С. 163–172.
Shokin, Yu.I., Fedotov, A.M. Library, working around the clock // ЕКО. 2000. No 6. P. 163–172. (in Russ.)
- [20] **Шокин Ю.И., Федотов А.М.** Информационная система Сибирского отделения РАН // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Сб. докл. Второй Всерос. науч. конф., Протвино, 26–28 сентября 2000 г. Протвино: ГНЦ ИФВЭ, 2000. С. 6–15.
Shokin, Yu.I., Fedotov, A.M. Information System SB RAS // Digital libraries: Advanced methods and technologies, digital collections: Coll. reports, Protvino, 26–28 September, 2000. Protvino: GNTs IFVE, 2000. P. 6–15. Available at: <http://web.ihep.su/library/pubs/aconf00/dconf00/ps/028.pdf> (in Russ.)
- [21] **Шокин Ю.И., Федотов А.М.** Электронная библиотека Сибирского отделения РАН // Информационно-библиотечное обеспечение науки. Проблемы интеграции информационных ресурсов: Сб. науч. тр. М.: Информ.-библ. совет РАН, 2000. С. 118–128.
Shokin, Yu.I., Fedotov, A.M. Electronic library of the Siberian Branch of the Russian Academy of Sciences // Information and Library Support of science. Problems of integration of information resources: Sat. sci. tr. Moscow: IBS RAS, 2000. P. 118–128. (in Russ.)
- [22] **Жижимов О.Л., Мазов Н.А., Федотов А.М.** Некоторые заметки об эволюции цифровых репозиторий традиционных библиотек к полнофункциональным электронным библиотекам // Вест. Владивосток. гос. ун-та экономики и сервиса. 2010. Т. 7, № 3. С. 55–63.
Zhizhimov, O.L., Mazov, N.A., Fedotov, A.M. Some notes on the evolution of digital repositories of traditional libraries for a full-featured digital libraries // Vestnik Vladivostokskogo Gos. Univ. Ekonomiki i Servisa. 2010. Vol. 7, No. 3. P. 55–63. (in Russ.)
- [23] **Антопольский А.Б., Вигурский К.В.** Концепция электронных библиотек // Электронные библиотеки. 1999. Т. 2, вып. 2. Адрес доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/antopol> (дата обращения: 04.05.2013).
Antopolskiy, A.B., Vigurskiy, K.V. Concept of digital libraries // Russian Digital Libraries Journal. 1999. Vol. 2, iss. 2. Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/1999/part2/antopol> (accessed: 04.05.2013). (in Russ.)
- [24] **Федотов А.М.** Концептуальные подходы к построению распределенных систем // Тр. Междунар. конф. по вычисл. математике (МКВМ-2004): Рабочие совещания. Новосибирск: ИВМиМГ СО РАН, 2004. С. 132–143.
Fedotov, A.M. Conceptual approach to building distributed systems // Proceedings of the International Conference on Computational Mathematics (ICCM-2004): Workshops. Novosibirsk ICMMG SB RAS, 2004. P. 132–143. (in Russ.)
- [25] **Шокин Ю.И., Федотов А.М.** Поддержка и развитие распределенных информационно-вычислительных ресурсов в СО РАН // Вестн. КазНУ им. аль-Фараби. Математика, механика, информатика. 2004. Т. 42, № 3. Ч. 4. С. 324–334.
Shokin, Yu.I., Fedotov, A.M. Support and development of distributed information resources of SB RAS // Vestnik Al-Farabi Kazakh. National University. Mathematics, Mechanics, Informatics. 2004. Vol. 42, No 3. Pt 4. P. 324–334. (in Russ.)
- [26] **Шокин Ю.И., Федотов А.М.** К вопросу о развитии информационной инфраструктуры СО РАН // Вычисл. технологии. 2009. Т. 14, № 6. С. 127–137.

- Shokin, Yu.I., Fedotov, A.M.** About information infrastructure SB RAS // Computational Technologies. 2009. Vol. 14, No 6. P. 127–137. (in Russ.)
- [27] **Земсков А.И., Шрайберг Я.Л.** Электронные библиотеки: учеб. пособие. 3-е изд. М.: ГПНТБ России, 2004.
Zemskov, A.I., Shraiberg, Y.L. Digital libraries. Moscow: National Public Library for Science and Technology, 2004. 130 p. (in Russ.)
- [28] **Воройский Ф.С.** Электронные и традиционные библиотеки суть не одно и то же // Электронные библиотеки. 2003. Т. 6, № 5. Адрес доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part5/voroisky> (дата обращения: 04.05.2010).
Voroisky, F.S. Traditional and digital libraries are not one and the same // Russian Digital Libraries Journal. 2003. Vol. 6, No. 5. Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part5/voroisky> (date of access: 04.05.2010). (in Russ.)
- [29] **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 // Электронные библиотеки. 2005. Т. 8, № 1. Адрес доступа: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part1/AEEKFH>
Akimov, S.I., Elizarov, A.M., Ershova, T.V., Kogalovskiy, M.R., Fedorov, A.O., Hohlov, Yu.E. Research and methodological support to the development of scientific digital libraries // Russian Digital Libraries Journal. 2003. Vol. 8, No. 1. Available at: <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part1/AEEKFH> (in Russ.)
- [30] **Вигурский К.В.** Что такое электронная библиотека? // Докл. конф. “Информационные технологии в образовании — 2005”. Адрес доступа: <http://rd.feb-web.ru/library.htm>.
Vigurskiy, K.V. What is a digital library? // Reports of conf. “Information Technologies in Education — 2005”. Available at: <http://rd.feb-web.ru/library.htm> (in Russ.)
- [31] **Федотов А.М., Баракхин В.Б., Жижимов О.Л., Федотова О.А.** Технология создания корпоративных информационных систем учета трудов научных работников // Вест. НГУ. Информационные технологии. 2011. Т. 9, №. 2. С. 31–41.
Fedotov, A.M., Barakhnin, V.B., Zhizhimov, O.L., Fedotova, O.A. The technology of creation of corporate information systems for accounting resources, created by researchers // Vestnik NSU. Information Technologies. 2011. Vol. 9, No. 2. P. 31–41. (in Russ.)
- [32] **Жижимов О.Л., Федотов А.М., Федотова О.А.** Построение типовой модели информационной системы для работы с документами по научному наследию // Вест. НГУ. Информационные технологии. 2012. Т. 10, № 3. С. 5–14.
Zhizhimov, O.L., Fedotov, A.M., Fedotova, O.A. Building a generic model of information system for working with documents on the scientific heritage // Vestnik NSU. Information Technologies. 2012, Vol. 10, No. 2. P. 5–14. (in Russ.)
- [33] **Резниченко В.А., Проскудина Г.Ю., Кудим К.А.** Концептуальная модель электронной библиотеки // Тр. XI Всерос. науч. конф. RCDL-2009. Петрозаводск, Карелия, 17–21 сентября 2009 г. Петрозаводск: КНЦ РАН. 2009. С. 23–31.
Reznichenko, V.A., Proskudina, G.Yu., Kudim, K.A. Conceptual model of digital library // Proc. of the XI All-Russian Res. Conf. RCDL-2009. Petrozavodsk: KNTS RAN, 2009. P. 23–31. (in Russ.)
- [34] ISO-14721 Reference Model for an Open Archival Information System (OAIS): Recommended Practicle: CCSDS 650.0-M-2 (Magenta Book). June 2012. Available at: <http://public.ccsds.org/publications/archive/650x0m2.pdf>
- [35] **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.
- [36] **Шокин Ю.И., Федотов А.М., Гуськов А.Е., Жижимов О.Л., Столяров С.В.** Электронные библиотеки — путь интеграции информационных ресурсов Сибирского отделения РАН // Вест. КазНУ. Математика, механика, информатика. 2005. № 2. С. 115–127.
Shokin, Yu.I., Fedotov, A.M., Guskov, A.E., Zhizhimov, O.L., Stolyarov, S.V. Digital libraries — the path of integration of information resources of the SBRAS // Vestnik KazSU. 2005. No. 2. P. 115–127. (in Russ.)
- [37] **Otlet, P.** Traite de documentation. Bruxelles: Ed. Mundaneum, 1934.
- [38] ГОСТ Р ИСО/МЭК ТО 10000-2-99. Информационная технология. Основы и таксономия функциональных стандартов. Ч. 2. Принципы и таксономия профилей ВОС. Адрес доступа: <http://vsegost.com/Catalog/38/38074.shtml>
GOST R ISO/IEC TR 10000-2:1999. Information technology. Framework and taxonomy of International Standardized Profiles. Pt 2. Principles and taxonomy for OSI profiles. Available at: <http://vsegost.com/Catalog/38/38074.shtml>
- [39] **Бездушный А.Н., Бездушный А.А., Серебряков В.А., Филиппов В.И.** Интеграция метаданных Единого Научного Информационного Пространства РАН. М.: Вычисл. центр им. А.А. Дородницына РАН, 2006. 238 p.
Bezdushnyu, A.N., Bezdushnyu, A.A., Serebryakov, V.A., Filippov, V.I. Integration of metadata of the Unified Scientific Information Space of the Russian Academy of Sciences. Moscow: Dorodnicyn Computing Centre of RAS, 2006. 238 p. (in Russ.)
- [40] **Федотов А.М., Баракхин В.Б., Жижимов О.Л., Федотова О.А.** Модель информационной системы для поддержки научно-педагогической деятельности // Вест. НГУ. Информационные технологии. 2014. Т. 12, № 1. С. 89–101.
Fedotov, A.M., Barakhnin, V. B., Zhizhimov, O.L., Fedotova, O.A. A model of information system to support scientific and educational activities // Vestnik NSU. Information Technologies. 2014. Vol 12, No. 1. P. 89–101. (in Russ.)
- [41] **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 // Вест. НГУ. Информационные технологии. 2011. Т. 9, № 1. С. 89–101.
Fedotov, A.M., Zhizhimov, O.L., Knyazeva, A.A., Kolobov, O.S., Mazov, N.A., Turchanovsky, I.Yu., Fedotova, O.A. Problems of authority control for distributed digital libraries and bibliographic database // Vestnik NSU. Information Technologies. 2011. Vol. 9, No. 1. P. 89–101. (in Russ.)
- [42] **Захаров А.А., Серебряков В.А.** Система управления электронными библиотеками LibMeta // Тр. XII Всерос. науч. конф. RCDL-2010. Казань: Казан. ун-т, 2010. С. 28–37.
Zakharov, A.A., Serebryakov, V.A. Digital library management system LibMeta // Proceedings of the RCDL-2010. Kazan: KSU, 2010. P. 28–37. (in Russ.)
- [43] DCM I — Dublin Core Metadata Initiative. Available at: <http://www.dublincore.org/>
- [44] DSpace: an open source solution for accessing, managing and preserving scholarly works / MIT Libraries, HP Labs. 2007. Available at: <http://www.dspace.org/>
- [45] EPrints Free Software. EPrints for Digital Repositories / School of Electronics and Computer Science, University of Southampton, UK, 2008. Available at: <http://www.eprints.org/>
- [46] Fedora Repository System / Fedora Commons: Gordon and Betty Moor Foundation, Cornell University Information Science, University of Virginia Library, The Andrew W. Mellon Foundation, 2007. Available at: <http://www.fedora-commons.org/>

- [47] Кудим К.А., Проскудина Г.Ю., Резниченко В.А. Сравнение систем электронных библиотек EPrints 3.0 и DSpace 1.4.1 // Тр. IX Всерос. науч. конф. RCDL-2007, Переславль-Залесский, 15–18 октября 2007 г. Переславль-Залесский: Изд-во “Университет города Переславля”, 2007. С. 241–252.
Kudim, K.A., Proskudina, G.Yu., Reznichenko, V.A. Comparison of repository systems EPrints 3.0 and DSpace 1.4.1 // Proceedings of the RCDL-2007. Pereslavl-Zalesskiy: Universitet goroda Pereslavlya, 2007. P. 241–252. (in Russ.)
- [48] The Open Archives Initiative Protocol for Metadata Harvesting: Protocol Version 2.0 of 2002-06-14 / The OAI Executive; OAI Technical Committee. 2004. Available at: <http://www.openarchives.org/>
- [49] RFC 4510: Lightweight Directory Access Protocol (LDAP): Technical Specification Road Map / OpenLDAP Foundation. 2006. Available at: <http://www.apps.ietf.org/rfc/rfc4510.html>
- [50] ANSI/NISO Z39.50–2003. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification. NISO Press, Bethesda, Maryland, USA, Nov., 2002.
- [51] SRU (Search/Retrieve via URL). Available at: <http://www.loc.gov/standards/sru/> (accessed: 23.08.2013).
- [52] 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.

Поступила в редакцию 25 сентября 2015 г.

Technologies for designing of distributed information systems to support research

SHOKIN, YURIY I., FEDOTOV, ANATOLIY M.*, ZHIZHIMOV, OLEG L.

Institute of Computational Technologies SB RAS, Novosibirsk, 630090, Russia

*Corresponding author: Fedotov, Anatoliy M., e-mail: fedotov@sbras.ru

Purpose. This article addresses the history and description of technological approaches used for the design of distributed information systems based on the example of the Siberian Branch of the Russian Academy of Sciences. It describes the architectural solutions developed for the establishment of information systems designed to support research and for the principles of its integration with the external sources.

Methodology. We define the functional requirements to the design such an information system that provides:

- Uniform distributed information environment.
- Information support aimed at research activities in both fundamental and applied areas.
- Support for professionally-oriented training systems and an exchange of scientific documents.
- Support for professionally-oriented access systems and interfaces for data warehouses.
- Collective use of the acquired electronic literature, catalogs, bibliographic databases and publications.
- Support for electronic versions of scientific journals published by the scientific organization.

— Support for the adoption and implementation of organizational and administrative decisions.

These requirements are determined, at first, by the information needs of researchers, and, secondly, ensuring reliability and long-term storage of information. The principles of organization of such a system is also defined.

Findings. We discuss the rules of representation and transformation of metadata and describe the work with dictionaries, which are used to organize and classify information resources and to model the relationships between them.

Originality/value. Finally, examples of practical implementation of these technologies are described using the existing information systems.

Keywords: information system, digital library, dictionary catalog, distributed information resources, classification of information resources, data integration, digital repository, research, information retrieval, information retrieval thesaurus, key terms, metadata, bibliographic databases.

Received 25 September 2015