

Межбиблиотечная кооперация в электронный век – опыт АРБИКОН

Племнек А.И., к.т.н.,

исполнительный директор,

Ассоциация региональных библиотечных консорциумов

Современное состояние в информационном обеспечении процессов науки, образования и культуры можно охарактеризовать в целом как бурно и несогласованно развивающиеся. Видимо, в целом, это можно обосновать высокими темпами развития ИКТ и экспоненциальным ростом информационных источников. Наблюдая за этими процессами как в мире, так и стране, можно сделать некоторые выводы по тенденциям развития систем информационного обеспечения, а именно отметим динамику перехода:

- от централизованных систем (с централизованным хранилищем информационных ресурсов) к распределенным;
- от однородных систем к гетерогенным, с точки зрения программно-аппаратного обеспечения узлов распределенной системы;
- от объединения однородных информационных ресурсов (ИР) к объединению разнородных (электронных и традиционных; по различным отраслям знаний; образовательных, библиотечных, музейных и пр.).

В целом, эти тенденции можно обозначить как построение единого распределенного информационного пространства страны, совместимого с мировым информационным пространством. Использование Интернет в качестве транспортной среды и средства доступа к ИР является уже фактом, который не оспаривается.

Примерно по указанным выше принципам происходило развитие распределенной библиотечно-информационной системы членов Ассоциации региональных библиотечных консорциумов (АРБИКОН), созданной в 2002 году. В конце 19990-х годов в России на инициативной основе была начата серия проектов по объединению библиотек в консорциумы для повышения эффективности процесса информационного обеспечения населения. Идея объединения была достаточно прозрачна: библиотека предоставляла свои ресурсы, получая при этом возможность использовать ресурсы других библиотек. Для технической реализации такого обмена информации требовалось разработать технологию, обеспечивающую совместимость на уровне протоколов и форматов обмена данными, оставляя при этом библиотеке на локальном уровне свободу выбора собственного решения для автоматизации.

Особенностью процесса создания единого распределенного информационно-библиотечного пространства является то, что изначально технологии развивались на основе общей концепции построения многоуровневого иерархического информационного пространства, базирующейся на национальных и международных стандартах.

К АРБИКОН объединяет около более 200 российских библиотек, среди которых:

- 75 массовых библиотек,
- 83 библиотеки ВУЗов,
- 12 библиотек сети РАН,
- ведомственные и прочие библиотеки

К 2006 году создана инфраструктура, объединяющая электронные ресурсы библиотек 57 регионов страны в единой распределенной информационной системе. Инфраструктура обеспечивает взаимодействие по протоколу Z39.50, международному стандарту (ISO), определяющему функции поиска извлечения, заказа и другие дополнительные функции в распределенной библиотечной системе; для библиографических описаний выбран национальный формат RUSMARC.

Общее количество электронных ресурсов в распределенной библиотечной информационной системе, объединяющей членов АРБИКОН, по состоянию на конец 2005 года составляет:

- 38 млн. библиографических записей
- 500 тыс. полнотекстовых электронных ресурсов.

Этот объем существенно превышает фонды крупнейших национальных и федеральных библиотек страны.

Развитие служб в распределенной информационной системе является инновационным процессом, сложность которого превышает сложность создания и сопровождения централизованной информационной системы. Однако только такой подход обеспечивает актуальность данных (используются локальные данные, не выполняется дополнительное перемещение данных, проводящее к неминуемым задержкам в обновлении информации) и оперативность выполнения сервисов (заказ на выполнение сервиса поступает непосредственно к держателю ресурса). Система в силу заложенных в ее основу решений является масштабируемой, настраиваемой и развиваемой. Имеющаяся базовая сеть может быть использована как для решения общих задач информационного обеспечения населения в области науки, культуры и образования, так и для решения более узких задач по информационному обеспечению определенной отрасли знаний или категории населения путем специализации сети на предоставление определенной категории ресурсов и определенного вида сервисов. У АРБИКОНа есть опыт создания таких систем: медицинский портал, объединяющий ресурсы только медицинских библиотек и предоставляющий специализированные средства навигации (тезаурус по медицине) для поиска ресурсов; музейно-библиотечный портал, обеспечивающий одновременно доступ к музейным экспонатам и информационным ресурсам по этой тематике, хранящимся в библиотеках; портал библиотек для слепых.

За годы своего существования АРБИКОН продемонстрировал свою устойчивость и стабильность развития. В настоящее время приоритетными направлениями развития являются повышение качества и ассортимента предоставляемых информационных ресурсов и сервисов (поиска ИР, заказа и доставки электронной копии, т.д.). Особое внимание уделяется наполнению распределенной информационной системы качественным контентом – электронными ресурсами для нужд образования и науки. Особую роль в этом процессе играют университеты, в которых создаются электронные библиотеки, объединяющие создаваемые в организации электронные ресурсы: учебники, учебные пособия, авторефераты диссертаций и диссертации, отсканированные материалы из фондов библиотеки, материалы конференций и прочее. Инфраструктура АРБИКОН объединяет эти коллекции в единое виртуальное целое. Ведется разработка механизмов лингвистического анализа полных текстов электронных документов для повышения релевантности поиска полнотекстовых электронных ресурсов, совмещения средств поиска по структурированным и неструктурированным данным. Разработаны типовые решения по соблюдению норм авторского права, когда соответствующие соглашения на размещение ресурса заключаются с правообладателем.

Знаменательными являются решения, принятые на III съезде АРБИКОН в июле 2006 г. Решение съезда был утвержден переход к проектной деятельности, когда участники определяют тот набор направлений работы АРБИКОН, в котором они планируют участвовать, создавая ресурсы и/или предоставляя сервисы. Но при этом все члены АРБИКОН должны удовлетворять базовым условиям, определяющим доступность электронных каталогов в объединенном информационном пространстве партнерства. Введение проектов АРБИКОН позволяет, таким образом, каждой библиотеке выбрать удовлетворяющие именно ее условия сотрудничества с партнерами. При этом общий принцип деятельности в проектах остается прежним – предоставляя собственные ресурсы, библиотека получает доступ к ресурсам всех других участников. Сейчас в процессе формирования следующие проекты АРБИКОН:

- межрегиональной аналитической росписи журнальных статей (МАРС), создающий информационный ресурс аналитической росписи около 1200 российских журналов;

- электронного заказа и доставки документов (ЭДД) на основе сводного каталога журналов;
- распределенной электронной библиотеки.

На наш взгляд, устойчивость распределенной библиотечной системы определяется логикой формирования ее инфраструктуры: она основывается на библиотеках, миссией которых традиционно является именно информационное обеспечение. Другим фактором успеха явилось довольно жесткое требование следования стандартом. В середине 90-х в стране не было ни одного сервера Z39.50, сейчас практически все библиотечные системы поддерживают его. Отметим, что протокол Z39.50 признается во всем мире как одно из самых мощных интегрирующих средств, но и одно из самых сложных в реализации.

Дальнейшее развитие библиотечной сети видится в интеграции с другими информационными системами, и проекты по объединению Интернет систем и библиотечных технологий становятся характерной чертой мирового информационного пространства. Для этого требуется гармонизировать протокол Z39.50 с протоколами поиска в Интернет, формат RUSMARC для библиографических описаний с элементами Dublin Core для описания метаданных, гармонизация с XML-технологиями.

Идея такого развития была реализована, например, в проекте "Формирование открытых баз данных информационных ресурсов в области образования, науки и культуры", выполненном по заказу Минэкономразвития России в 2005 году. Участниками проекта стали следующие 5 организаций:

1. Всероссийский институт научной и технической информации Российской академии наук (ВИНИТИ РАН);
2. ФГУП НТЦ «Информрегистр»;
3. Некоммерческое партнерство «Ассоциация региональных библиотечных консорциумов (АРБИКОН);
4. Межрегиональная общественная организация в поддержку Программы ЮНЕСКО «Информация для всех»;
5. Государственное образовательное учреждение высшего профессионального образования Государственный университет – Высшая школа экономики.

Таким образом, в проекте участвовали представители госсектора и общественные организации, представители различных министерств и ведомств. Объединение столь разных организаций в едином консорциуме исполнителей проекта целесообразно по нескольким причинам, среди которых выделим лишь две основные. Во-первых, все они обладают существенными массивами публичных информационных ресурсов, и, во-вторых, вместе они имеют возможность комплексно подойти к решению задач проекта, формируя правовой, организационный и технологический базисы для создания единого интегрированного информационного пространства разнородных информационных ресурсов. Целью проекта являлось разработка Концепции единого пространства открытых баз данных и разработка прототипа, демонстрирующего работоспособность принятых решений. Отметим, что Концепция не направлена на создание новых хранилищ информационных ресурсов, копирующих и дублирующих данные уже существующих систем информационного обеспечения организаций и ведомств. Концепция направлена на объединение уже существующих систем информационного обеспечения в области образования, науки и культуры в единое виртуальное информационное пространство. Таким образом, создание единой системы открытых баз данных не заменяет другие, специализированные информационные системы в области образования, науки и культуры, а предоставляет новое качество за счет их объединения для обеспечения поиска в едином интегрированном пространстве. При этом разработанная технология не повторяет алгоритмы работы поисковых машин Интернет. Отличительной особенностью предлагаемого решения является то, что объединяемые в системе открытых баз данных ресурсы, как правило, размещены во внутренних базах данных организаций (в так

называемом «глубоком Интернете») и недоступны роботам поисковых машин для сканирования.

Следует сказать, что созданию подобных систем уделяется значительное внимание в развитых странах мира, как неотъемлемой черте электронного государства, прозрачной государственной власти, гарантирующей предоставление населению равных прав на доступ к информации. Вспомним в этой связи рекомендации комитета OASIS (Organization for the Advancement of Structured Information Standards, <http://www.oasis-open.org>) - некоммерческого международного консорциума. OASIS выполняет работу по развитию, согласованию и адаптации стандартов в области e-business, в общественном секторе. Консорциум основан в 1993 г., в его состав входят более 5000 участников из 600 организаций. В 2004 г. OASIS разработал рекомендации по обеспечению совместимости служб информационного поиска в электронном государстве. В докладе правительству США констатируется, что государственной власти рекомендуется обеспечить интероперабельность в распределенных информационных системах путем принятия единого (общего) поискового сервиса. Поисковый сервис должен базироваться на международном стандарте ISO 23950, обладающим высоким уровнем интероперабельности в различных профессиональных сообществах, для различных типов данных и держателей информационных ресурсов. Государство должно обеспечивать этот поисковый сервис дополнительно к другим поисковым механизмам, которые могут быть использованы для других целей, не связанных с обеспечением широкомасштабной совместимости.

Основными принципами технологического обеспечения системы при формировании единого информационного пространства в рамках проекта Минэкономразвития России, являются распределенность, независимость и неоднородность. Система является распределенной - состоящей из отдельных узлов, объединенных организационно и технологически в единое информационное пространство. В узлах размещаются различные (в общем случае независимые и неоднородные) информационные ресурсы и программно-технологические средства, взаимодействие которых при выполнении поиска следует обеспечить в едином информационном пространстве.

Для обеспечения интероперабельности узлов системы требуется формирование общесистемного профиля стандартов, включающего стандарты представления данных (открытых баз данных), стандарты запросов и стандарты передачи данных. Для представления данных используется стандарт метаданных Дублинского ядра (Dublin Core) - простой, но эффективный набор элементов для описания широкого спектра сетевых ресурсов. Для организации взаимодействия узлов используется сервисная архитектура, базирующаяся на технологии Web-сервисов (WSDL – Web Services Description Language) и стандартные протоколы SRW/SRU (Search/Retrieve Web Service, Search/Retrieve via URL).

Все узлы распределенной системы разделены на три уровня: глобальный, корпоративный и локальный.

Открытый информационный ресурс (базы данных, публикации, звуковые и видео файлы, пр.) различных организаций и ведомств (органов государственной власти, архивов, библиотек, музеев, научных и образовательных учреждений и пр.) располагается в узлах распределенной инфраструктуры системы на локальном уровне. Локальные информационные системы имеют различную реализацию – от страниц Web-сайта организации до сложных многоуровневых систем.

Корпоративный уровень является промежуточным между локальными узлами и единой системой навигации, этот уровень используется не всеми узлами локального уровня, т.е. является необязательным. Корпоративный узел объединяет локальные узлы по корпоративным правилам взаимодействия (корпоративный профиль), отличающимся от общесистемного профиля (например, архивный, библиотечный и т.п. профили). На

локальном или корпоративном уровнях обеспечиваются преобразование структуры данных каждого узла в общесистемную схему метаданных, а также доступность информационных ресурсов для поиска по стандартному протоколу.

На глобальном уровне размещаются информационные ресурсы, использование которых обеспечивает целостность распределенного информационного пространства как единой системы. К таким информационным ресурсам относятся:

- лингвистическое обеспечение (классификаторы, рубрикаторы, пр.), являющееся обязательным для выполнения поиска по единому информационному пространству. Такое лингвистическое обеспечение должно быть универсальным, актуальным и открытым. Лингвистическое обеспечение должно быть использовано во всех узлах сети или непосредственно, или опосредованно, путем однозначного отображения в другое лингвистическое обеспечение, используемое в узлах;
- реестры, выполняющие функцию государственной регистрации информационных ресурсов определенного уровня достоверности;
- эталонные информационные ресурсы, создаваемые по определенному регламенту и используемые затем для устранения дублирования работ различными организациями (ресурсы официальной документации, архивных документов, государственной библиографии и другие, создаваемые, как правило, по государственному заказу).

Общесистемные ресурсы глобального уровня представлены в едином пространстве в качестве Web-сервисов.

В едином пространстве создаются сервисы, обеспечивающие эффективное взаимодействие узлов: регистрации, адресно-справочные, лингвистические и пр.

На глобальном уровне располагается порталный механизм, единая система навигации, обеспечивающей доступ ко всему информационному пространству. Единая система навигации, выполняющая поиск по массиву информационных ресурсов всего единого виртуального пространства по единым универсальным правилам (общесистемный профиль стандартов). При этом пользователю становится доступна информация о местоположении информационного ресурса и сам информационный ресурс, в случае наличия у пользователей соответствующих полномочий на использование ресурса. Общая архитектура системы приведена на рисунке 1.

Система должна обеспечивать навигационные и спецификационные методы поиска. Навигационные методы предполагают навигацию по некоторым универсальным рубрикаторам для конкретизации предметной области поиска. Спецификационные методы обеспечивают формирование поискового запроса в виде набора атрибутов, в виде набора ключевых слов или в виде фразы-вопроса на естественном языке.

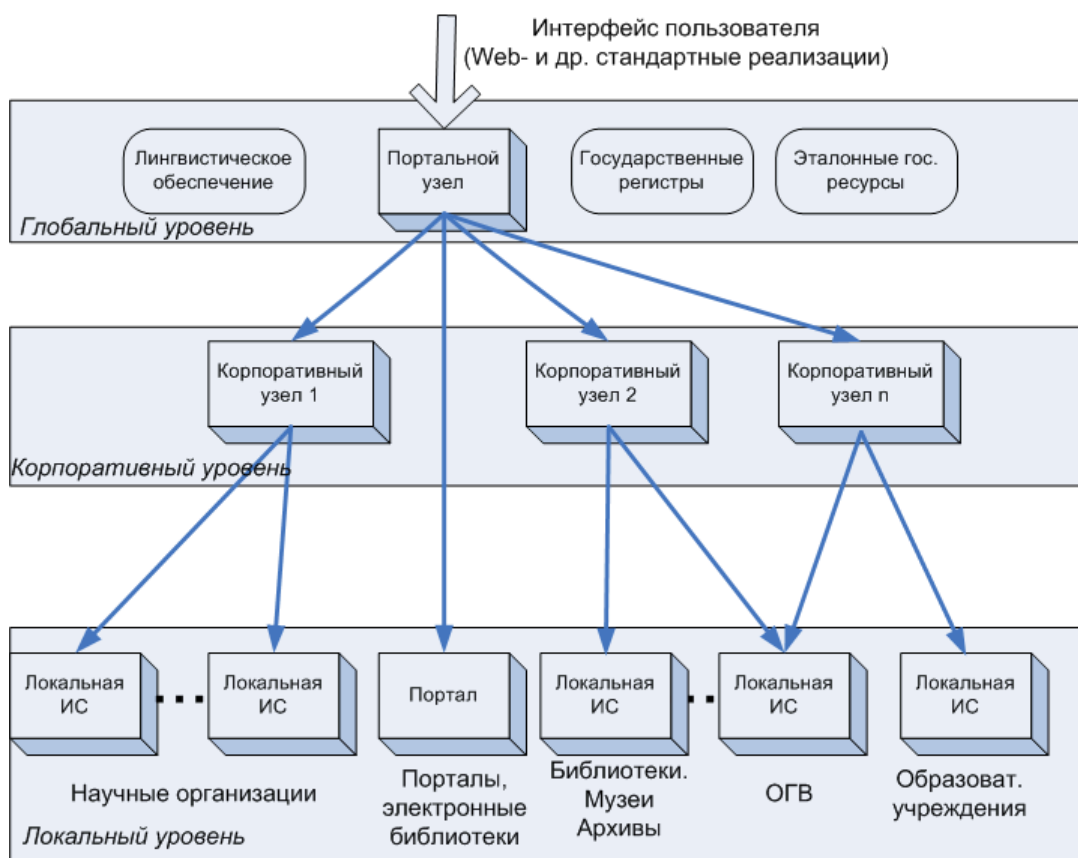


Рис.1. Общая архитектура распределенной системы открытых баз данных.

При разработке прототипа целевой системы открытых баз данных было выполнено объединение в распределенной информационной системе разнородных информационных ресурсов, а именно:

- библиографических записей библиотек – членов АРБИКОН;
- реферативных баз данных ВИНТИ;
- баз данных реестра «Информрегистра»;
- БД ГАСНТИ
- правовых БД
- электронных полнотекстовых ресурсов университетов.

Участниками разработан проект стандарта системы открытых баз данных, предполагающий использование:

- для представления информационных объектов - Dublin Core Metadata element set ISO Standard 15836:2003;
- для распределенного поиска - SRU (Search/Retrieve via URL);
- для формирования интерфейса пользователя - Hypertext Markup Language (HTML) v4.01, Стандарт W3C; Hypertext Transfer Protocol (HTTP) v1.1, Стандарт IETF; Cascading Style Sheets Language, level 2 revision 1 (CSS2.1), Стандарт W3C;
- для описания интерфейсов системного взаимодействия - Web Services Description Language (WSDL) v1.1;
- для службы каталогов - Lightweight Directory Access Protocol (LDAP) v3.

Для набора DC определен перечень обязательных полей и таблицы отображений этих элементов в атрибуты описаний ресурсов, входящих в систему открытых баз данных. АРБИКОН обеспечивал корпоративный уровень для библиотек при подключении их к системе открытых баз данных. АРБИКОН и ВИНТИ реализовали два варианта реализации портального уровня для всей системы.

Таким образом, впервые в стране был реализован прототип системы, выполняющей распределенный информационный поиск по разнородным ресурсам. Однако проект

Минэкономразвития показал, что основным препятствием для реализации такой системы в полном масштабе являются вопросы нормативно-правового обеспечения. Решению именно этих вопросов будет отдан приоритет в ближайшее время. Тем не менее, результаты проекта могут быть широко использованы уже сейчас. АРБИКОН видит их применение в совершенствовании информационных сервисов Партнерства, когда в едином пространстве могут быть объединены библиографические, полнотекстовые, мультимедийные, музейные и прочие ресурсы. Поскольку протокол SRU обладает меньшей функциональностью, чем Z39.50 (по сути, он включает только функции поиска - Search и извлечения – Retrieve из протокола Z39.50), то он не заменяет протокол Z39.50 в библиотечной сети, где актуально также обеспечение функцией заказа и других, специфических для библиотек, и поддерживаемых в Z39.50. Однако применение протокола SRU для интеграции разнородных ресурсов является более быстрым и дешевым решением, способным обеспечить подключение БД в течение примерно недели путем написания программы-шлюза, обеспечивающей отображение SRU-запросов специфическую структуру конкретной базы данных.

Результаты, полученные в рамках проекта Минэкономразвития РФ, используются в развитии ресурсов и сервисов АРБИКОНа. Так, они самым непосредственным образом будут использованы в Проекте распределенной электронной библиотеки, которая формируется в настоящее время.

Таким образом, ассоциация библиотек динамично развивается, активно используя самые передовые технологии для совершенствования информационно-библиотечных сервисов партнерства.