

Послушайте, ребята,
Что вам расскажет дед.
Земля наша богата,
Порядка в ней лишь нет.

А. К. Толстой

Модель электронной библиотеки по научному наследию

ФЕДОТОВ А. М., ФЕДОТОВА О.А.

Институт вычислительных технологий СО РАН,
Государственная публичная научно-техническая библиотека СО РАН

ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Одним из основных результатов созидательной, социальной и интеллектуальной человеческой деятельности является создание и накопление информационных ресурсов с целью их дальнейшего использования и недопущения утраты опыта предыдущих поколений.

*Окружающий нас мир
непознаваем, ввиду того, что
мы изучаем не его, а лишь
наше представление о нем*

Эммануил Кант

ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Не будет преувеличением сказать, что уровень развития технологий накопления информации и эффективности использования накопленной ранее информации на протяжении всей истории человечества значительно влиял на уровень развития производительных сил. Утеря информации приводила к отбрасыванию цивилизации на века назад.

ИНФОРМАЦИОННЫЕ РЕСУРСЫ

Однако, чтобы эффективно пользоваться накопленной ранее информацией, необходимы специальные инструменты и специальные технологии, при помощи которых могут быть реализованы специальные приемы работы с информацией.

*Многие вещи нам не
понятны не потому, что
наши понятия слабы: но
потому, что сии вещи
не входят в круг наших
понятий*

Козьма Прутков

ПРОБЛЕМА ПОИСКА ИНФОРМАЦИИ

*Если Вам все равно, где Вы находитесь
значит Вы не заблудились*

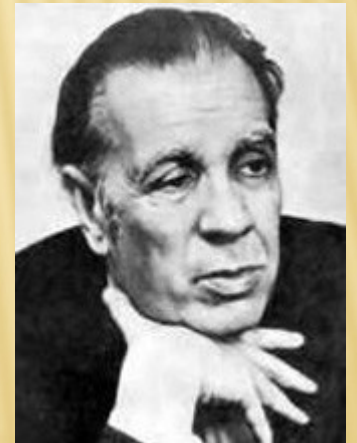
Одна из вечных проблем человеческого сообщества. На протяжении своего многотысячелетнего развития его представители неустанно находятся в поиске того, где находится что-либо: *пищи, жилища, пастбищ, дорог, сокровищ* и т. п.

Человечество постоянно находится в поиске знаний: «информации о том, где лежат сокровища».

ПРОБЛЕМА ПОИСКА ИНФОРМАЦИИ

Великий аргентинский писатель (библиотекарь) Хорхе Луис Борхес в своем эссе «Четыре цикла» писал, что в мировой литературе вечными являются четыре темы:

- Падение города
- Возвращение героя
- Поиск
- Самопожертвование бога.



ПРОБЛЕМА ПОИСКА ИНФОРМАЦИИ

«Есть правила для выбора решения, но нет правил для выбора этих правил»

Энон

- ✘ Нетрудно заметить, что наиболее часто встречающейся как в литературе, так и в реальности является третья тема - **поиск**, ибо четвертая тема выходит за рамки обычного человеческого опыта, а две первые проявляются лишь в «минуты мира роковые»
- ✘ «Вавилонская библиотека»

ПРОБЛЕМА ПОИСКА ИНФОРМАЦИИ

- ✘ Проблема поиска информации - одна из вечных проблем человечества. Чтобы решить проблему доступа к информации, человечество создало библиотеки - универсальную систему хранения, систематизации и каталогизации «информации и знаний».
- ✘ Александрийская библиотека

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

- ✘ Электронная библиотека (ЭБ) - это структурированная каталогизированная коллекция разнородных электронных документов, снабженная средствами навигации и поиска (в отличие от печатных изданий, микрофильмов и других носителей).
- ✘ ЭБ должна быть способна не только обеспечить многосторонний поиск в каталоге, но и предоставить пользователю непосредственно найденный ресурс (публикацию, документ, фотографию, описание факта и др.), а также дополнительные сведения о нем, например, информацию об авторах, библиографию, организации и т.п.

НАУЧНОЕ НАСЛЕДИЕ

- ✘ Научное наследие – это опубликованные результаты научных исследований и экспериментов, библиографические и фактографические базы данных, сведения об ученых, их научной деятельности, публикациях, проектах и т.п., а также большое количество неопубликованных документов, таких как отчеты, письма, воспоминания, записки, фотоматериалы и т.п.
- ✘ Эти ресурсы представляют большой интерес для научного сообщества и представителей общественности.

НАУЧНОЕ НАСЛЕДИЕ

Если вам непонятно какое-то слово в техническом тексте, не обращайтесь на него внимания. Текст полностью сохраняет смысл и без него.

Купер

- ✘ Однако в настоящий момент значительная часть информационных ресурсов по научному наследию недоступна широкому кругу научной общественности, а ресурсы, представленные в Интернет, разрознены, недостаточно систематизированы и структурированы.
- ✘ При создании их описаний недостаточное внимание уделяется вопросам интероперабельности – слабо применяются соглашения и рекомендации по стандартизации представления документов и средства интеграции разнородных информационных ресурсов.

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

Основными целями создания ЭБ по научному наследию являются:

- ✘ предоставление научным работникам быстрого доступа к информационным ресурсам по научному наследию;
- ✘ предоставление результатов фундаментальных научных исследований мировому сообществу;
- ✘ предотвращение утраты ценных научных коллекций для будущих поколений ученых;
- ✘ создание новых технологий научных исследований, эффективного инструментария для их проведения

НАУЧНОЕ НАСЛЕДИЕ

- ✘ Михайлов А.И., Черный А.И., Гиляревский Р.С. выделили два типа информационных потребностей исследователя: потребности в сведениях об источниках необходимой научной информации и потребности в **самой** необходимой научной информации (фактах).
- ✘ Для удовлетворения информационных потребностей первого предназначены информационные системы, получившие название «документальных» (библиографических), второго типа – «фактографических».
- ✘ В настоящее время наиболее востребованным средством информационного обеспечения научной деятельности становятся интеллектуальные системы, сочетающие возможности информационных систем обоих названных типов и позволяющие удовлетворять информационные потребности квалифицированного пользователя в соответствии со схемой «документ → факт → рассуждение».

НАУЧНОЕ НАСЛЕДИЕ

- ✘ Существует достаточно много технологических разработок информационных систем для электронных библиотек, так или иначе ориентированных на поддержку научных исследований.
- ✘ Среди них следует отметить информационные системы, близкие к фактографическим, например ИСИР (ЕНИП) РАН – ЭБ по научному наследию, ИРИС СО РАН, euroCRIS, и документальные, например eLibrary, Информика, MathNET и т.п.

<http://www.eurocris.org/>

<http://elibrary.ru/>

<http://www.informika.ru/>

<http://www.mathnet.ru/>

Предназначение информационных систем

Большинство людей не получают того, чего хотят, а всё потому, что сами не знают чего хотят.
Как исполнять желания, если их нет?
Дж.Б.Пристли. Тридцать первое июня.

Цель работы информационных систем – обеспечение конечного пользователя необходимой информацией.

Пользователя, как правило, не интересует, как устроена технологическая «кухня» информационной системы.

Более того, чем меньше эта «кухня» пользователю заметна, тем лучше построена та или иная информационная система.

Следует думать, что использование проверенных временем и практикой типовых решений позволит обеспечить эту технологическую «прозрачность».

ПРОБЛЕМЫ

Сложные проблемы всегда имеют простые, легкие для понимания неправильные решения

Закон Х.Л.Менкина

- ✘ модели и стандарты представления информации и метаинформации;
- ✘ автоматическая классификация информации;
- ✘ доступ к распределенным и разнородным коллекциям (интероперабельность, масштабируемость, обнаружение релевантной информации, интеграция метаинформации);
- ✘ интерфейсы пользователей, визуализация и анализ данных,
- ✘ вопросы интеллектуальной собственности;

ПРОБЛЕМЫ

Я не знаю причины, по которой нам не следует этого делать, но, возможно, позже мы придумаем какую-нибудь.

Марк Дэвисон

- ✘ Помимо проблемы интеграции первичных информационных ресурсов в единое информационное пространство, стоит проблема разработки основных принципов организации таких ресурсов с целью их эффективного использования

МЕТАДАННЫЕ

*...Из окна чудесный вид на будущий стадион,
вот только бульдозеры окончат работу
и будет осушено болото.*

Дж.Б.Пристли

- ✘ обеспечивают механизмы интеграции информационных ресурсов из разных источников сведениями о свойствах этих ресурсов;
- ✘ являются источниками сведений о свойствах и содержании информационных ресурсов для механизмов управления данными в информационных системах;
- ✘ являются источником информации, необходимой для осуществления реинжиниринга информационных систем;
- ✘ обеспечивают представление сведений о системе, ее информационных и других ресурсах для различных приложений и пользователей системы.

ИНТЕГРАЦИЯ

- ✘ При этом стандартизации должны подлежать:
 - протоколы и интерфейсы доступа к данным и их поиска;
 - схемы и форматы представления данных;
 - правила контроля доступа к данным;
 - правила индексации данных и анонсирования сервисов.

ИС ДОЛЖНА ВКЛЮЧАТЬ

Когда мы пытаемся вытащить что-нибудь одно, оказывается, что оно связано со всем остальным.

Закон Муира

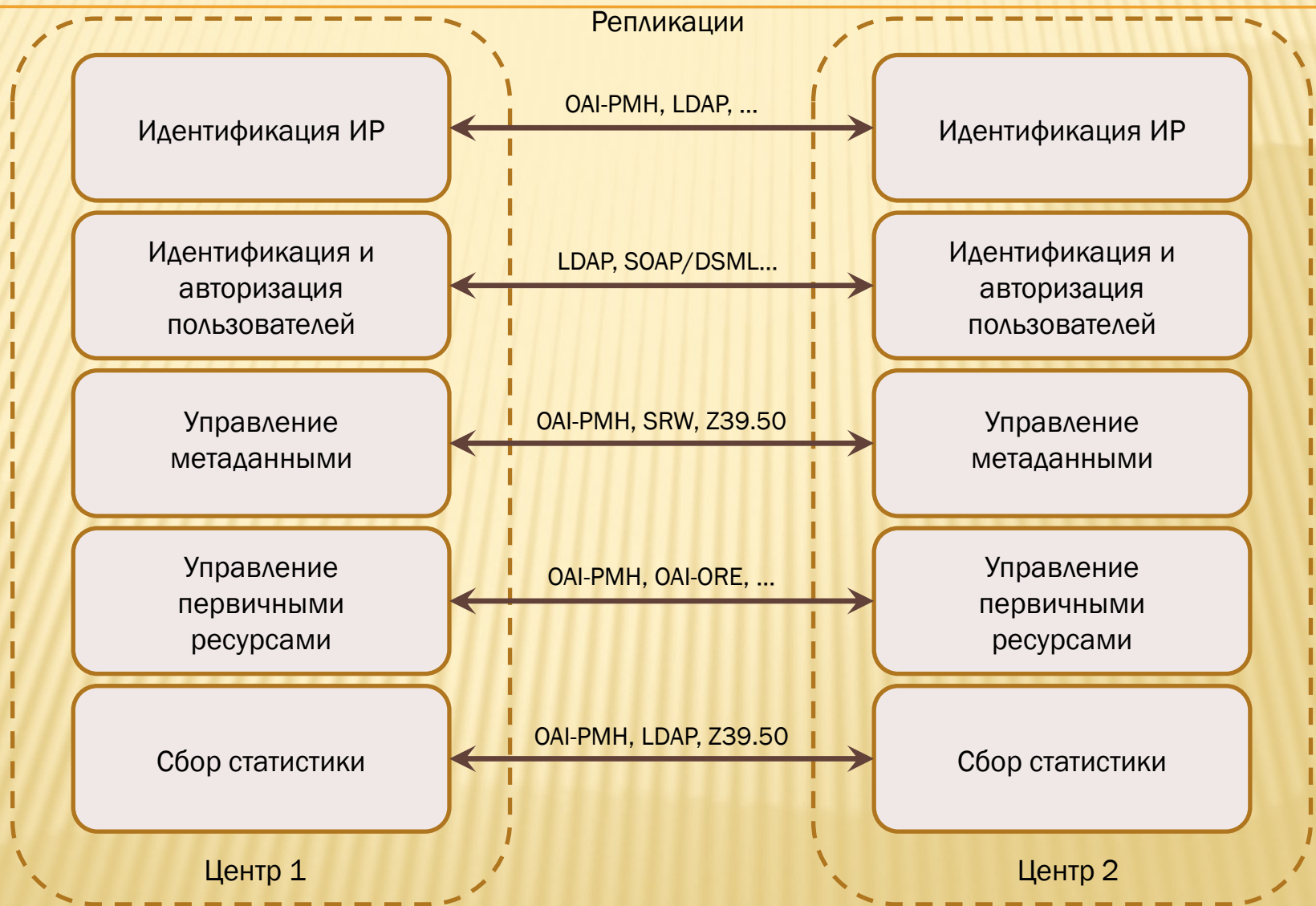
ИС должна содержать отдельные функциональные компоненты:

- ✘ Система идентификации информационных ресурсов
- ✘ Система идентификации, аутентификации и авторизации пользователей
- ✘ Система управления метаданными
- ✘ Система управления информационными ресурсами, в том числе полнотекстовыми
- ✘ Система сбора статистики
- ✘ Систему мониторинга доступности сервисов и ресурсов.

РЕАЛИЗАЦИЯ ИС

- ✘ 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.
- ✘ При этом в качестве протоколов сетевого взаимодействия должны выступать стандартные протоколы, например, OAI-PMH, Z39.50, SRW/SRU, LDAP и др.

ПОДСИСТЕМЫ ИНТЕГРИРОВАННОЙ СРЕДЫ



ПРОФИЛИ ИС И ПРОФИЛИ МЕТАДАННЫХ

- ✘ В настоящее время существуют достаточно мощные информационные системы, удовлетворяющие в той или иной степени потребности научных работников в информации, однако основной недостаток большинства систем – ограниченность возможностей обеспечения интеграции ресурсов как внутри каждой из систем, так и с внешними.
- ✘ Отметим, что основу разработки ИС составляют, прежде всего, стандарты и международные рекомендации, формирующие профиль ИС.

ПРОФИЛИ ИС И ПРОФИЛИ МЕТАДААННЫХ

- ✘ Информация в ИС отображает некоторые сущности (предметы, процессы, явления, персоны, публикации, факты, ключевые термины и т. п.), следует рассматривать информационную систему как множество информационных объектов – наборов данных, представляющих (описывающих) эти сущности в ИС.
- ✘ Эффективным средством описания информационных объектов в ИС являются метаданные – данные, являющиеся неотъемлемой частью информационного объекта и описывающие реальный объект или группу объектов.



ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

- ✘ Отличительной чертой электронной библиотеки является возможность параллельного использования различных поисковых механизмов и средств доступа к гетерогенным банкам электронных данных.
- ✘ Так как в качестве ответа на запрос к электронной библиотеке пользователю может быть представлен далеко не единственный электронный документ (или его фрагмент) и совсем не обязательно в текстовой форме, необходимы такие информационные системы, которые обеспечат эффективный комплексный поиск и анализ информации в коллекциях гетерогенных объектов.

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ

- ✘ 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 обеспечения 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.

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

Как бы кропотливо и тщательно вы ни готовили выборку, вам всегда могут сказать, что она неправильна и неприменима к данной проблеме.

Эдингтон

Исходя из целей ЭБ для поддержки научных исследований, можно сформулировать следующие функциональные требования к модели ЭБ:

- ✘ надежное долговременное и защищенное от исчезновения хранение информации;
- ✘ актуальность, полнота, достоверность происхождения документов;
- ✘ историчность информации;
- ✘ географическая привязка информации;
- ✘ наличие большого числа словарей-классификаторов (справочников), для обеспечения идентификации и классификации ресурсов;
- ✘ поддержка неоднородных и слабо структурированных информационных ресурсов;

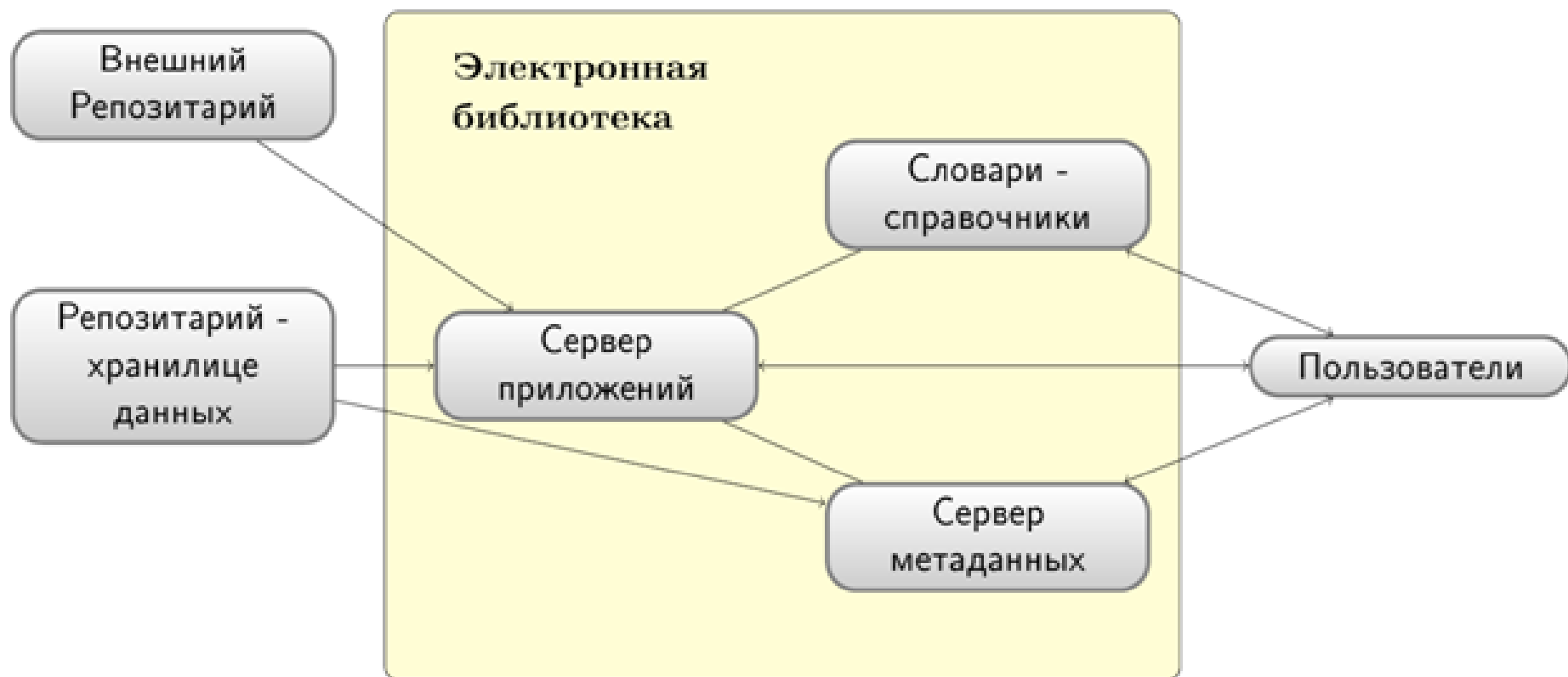
ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

Функциональные требования к модели ЭБ:

- ✘ поддержка взаимосвязей информационных ресурсов;
- ✘ предоставление информации пользователю в виде, выбранном пользователем;
- ✘ наличие интеллектуальных служб обслуживания запросов пользователя;
- ✘ наличие программных интерфейсов для поддержки аналитической работы пользователя с помощью программных приложений;
- ✘ поддержка требований интероперабельности как на программном, так и на семантическом уровне;
- ✘ поддержка работы с внешними источниками.

Наиболее важным выводом из вышесказанного является то, что информационная модель ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент: хранилище данных – репозиторий, сервер метаданных, сервер приложений, словари-справочники

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ



ВЫБОР МЕТАДААННЫХ

Эталонные модели ЭБ (RM):

- ✘ **CIDOC CRM** – музейные коллекции
- ✘ **FRBR** и **FRBRoo** – Международная федерация библиотечных ассоциаций и учреждений
- ✘ **DELOS DLRM**

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент: хранилище данных - репозитарий, сервер метаданных, сервер приложений, словари-справочники:

- ✘ Репозитарий (или депозитарий, далее ЦД) – это независимая система долговременного хранения и доступа к разнородным цифровым объектам, которая предназначена для предоставления электронных (цифровых) версий документов по научному наследию (книг, научных статей, репринтов, писем, изображений и других материалов, представленных в электронном виде).

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент: хранилище данных - репозитарий, сервер метаданных, сервер приложений, словари-справочники:

- ✘ Справочники – это набор баз данных, содержащих информацию об авторах и других персонах (авторитетные файлы), географических пунктах, городах, издательствах, имеющих отношение к конкретной научной школе, тематические словари-классификаторы, тезаурусы, описания предметной области данной научной школы и классификаторы документов в соответствии с МЕКОФ.

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ

ЭБ должна быть многоуровневой и состоять как минимум из следующих компонент: хранилище данных - репозитарий, сервер метаданных, сервер приложений, словари-справочники:

- ✘ Сервер метаданных должен обеспечить работу с метаданными – каталогизацию всех информационных ресурсов в соответствии с общепринятыми международными стандартами.
- ✘ Сервер приложений – основные сервисы ЭБ.

ХРАНИЛИЩЕ ДАННЫХ - РЕПОЗИТАРИЙ

- ✘ Для организации системы долговременного хранения информационных ресурсов (депозитариев) международной организацией по стандартизации (ISO) предложен стандарт ISO-14721 (OAIS – Open Archive Information System).
- ✘ Эталонная модель для стандарта OAIS – это концептуальная модель, которая была использована многими организациями для разработки наборов метаданных и организации крупных хранилищ цифровых объектов.
- ✘ Поддержка протоколов OAI, LDAP и SRW/SRU.

ХРАНИЛИЩЕ ДАННЫХ - РЕПОЗИТАРИЙ

- ✘ Наиболее популярные в мире системы ЦД, такие как **DSpace, E-Prints, Fedora, GREENSTONE** и др., строятся в рамках модели OAIS на основе технологий открытых систем, что дает принципиальную возможность их использования при построении информационных систем.
- ✘ Процесс интеграции ЦД в среду электронной библиотеки для этих систем отличается только в несущественных деталях и основан на модели агрегирования и распространения метаданных.
- ✘ Применение этой модели закреплено в протоколе OAI Protocol for Metadata Harvesting (далее OAI или OAI-PMH)

СЛОВАРИ

Словари (ключевые признаки, ключевые термины) – это особый вид метаданных, которые отражают наиболее существенные свойства объекта, имеющие наибольшее значение с точки зрения ИС, и их специфика определяется терминологией конкретной предметной области, которой посвящена ЭБ. Необходимо рассматривать различные типы ключевых терминов, а именно:

- + ключевые термины в стандартном понимании;
- + ключевые термины, описывающие персону;
- + ключевые термины, описывающие организацию;
- + ключевые термины, описывающие временные периоды;
- + ключевые термины, описывающие географические понятия,

СЛОВАРИ

- ✘ а также тематические словари-классификаторы, тезаурусы, описания предметной области данной научной школы и классификаторы документов в соответствии с МЕКОФ.
- ✘ Имеется ряд российских (например УДК, ГРНТИ) и международных (например ДКД, MSC2010, ORTELIUS) словарей для классификации научных данных.
 - + Однако в целом существующие словари содержат только общенаучную информацию и не годятся для систематизации и классификации материалов по научному наследию конкретной научной школы.

СЕРВЕР МЕТАДАННЫХ

- ✦ **Список необходимых сервисов для управления метаданными включает следующие:**
 - сервисы управления данными: сервисы каталогизации данных, т.е. создание вторичных информационных массивов, сервисы пакетной загрузки метаданных, сервисы заимствования метаданных из других информационных систем, сервисы синхронизации метаданных между разными информационными системами;
 - сервисы распределенного поиска в массивах разнородных метаданных;
 - сервисы извлечения метаданных в различных схемах и форматах;
 - сервисы просмотра индексов;
 - сервисы информирования о деталях конфигурации информационной системы и всех ее компонент;
 - сервисы контроля доступа к данным и метаданным;
 - сервисы предоставления доступа к контенту.

СЕРВЕР ПРИЛОЖЕНИЙ

- Обеспечивает сервисы необходимые для формирования информационных ресурсов с использованием и без использования диалоговых пользовательских интерфейсов.
- Сервисы позволяют использовать метаданные других информационных систем в диалоговом режиме и пакетных режимах.
- Их функциональность должна обеспечивать поиск и извлечение метаданных из других систем, конвертирование полученных метаданных в схемы и структуры локальной системы.

СЕРВЕР ПРИЛОЖЕНИЙ

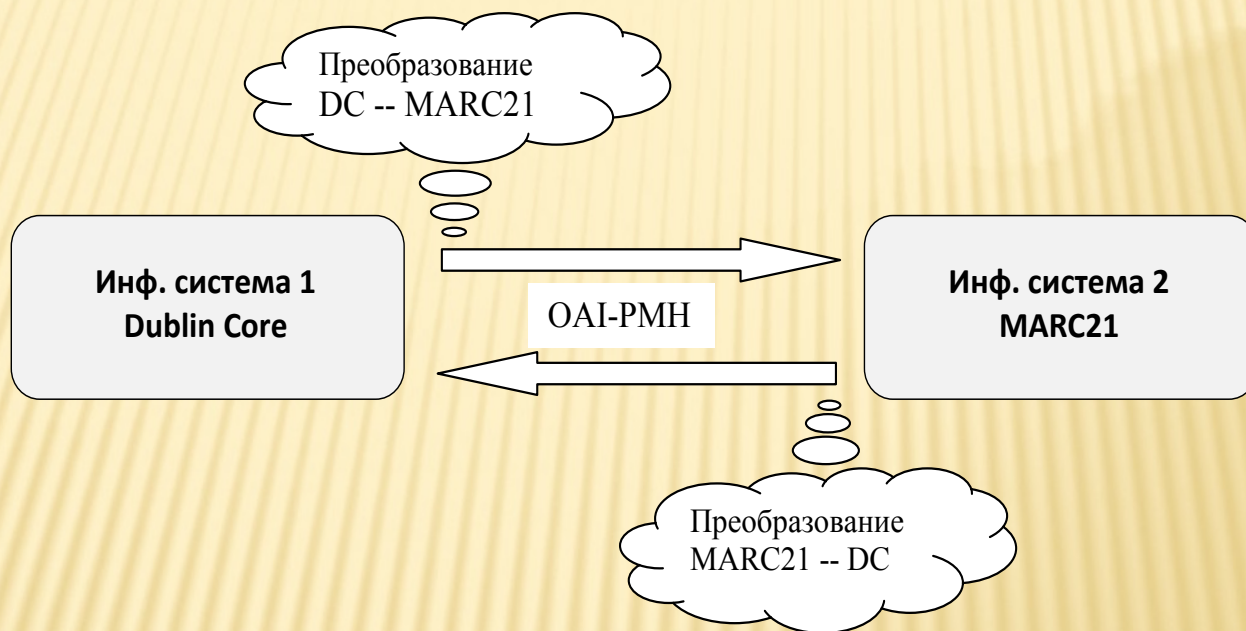
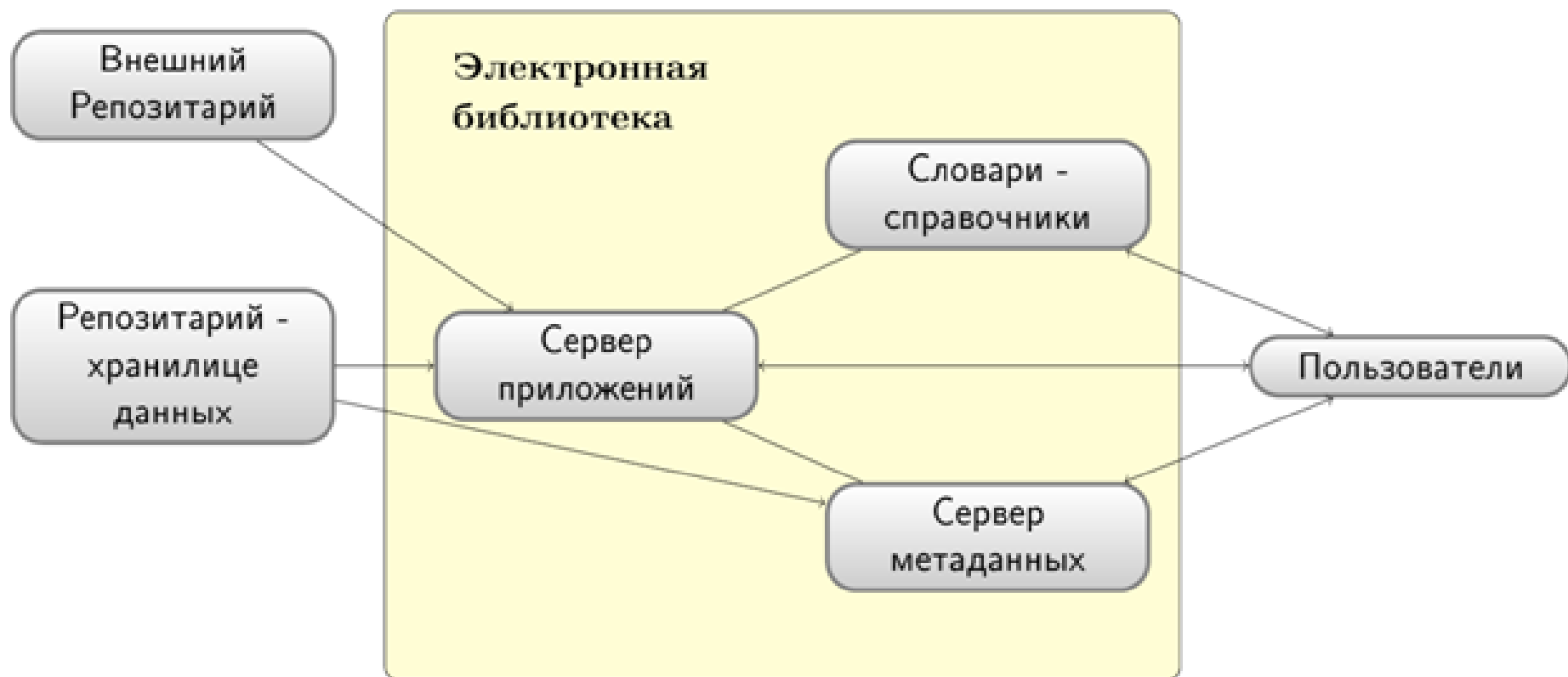


Схема обмена метаданными OAI-PMH

ЭБ НАУЧНЫХ ИССЛЕДОВАНИЙ



НАУЧНОЕ НАСЛЕДИЕ А.А.ЛЯПУНОВА

- ✘ Рассмотренная модель информационной системы, работающей с материалами научного наследия, реализуется на примере научной школы Алексея Андреевича Ляпунова – основателя теоретического программирования и российской (советской) кибернетики.
- ✘ Основной каталог информационных ресурсов сервера метаданных информационной системы строится в соответствии со схемой метаданных МЕКОФ и схемы метаданных Dublin Core. Для долговременного хранения документов использовался репозиторий DSpace. Нами была расширена стандартная схема метаданных DSpace полями, включающими основные требования МЕКОФ.
- ✘ Для поддержки процесса наполнения полнотекстовых баз, созданные профили метаданных были зарегистрированы в системе DSpace и в соответствии с ними были настроены рабочие процессы и пользовательский интерфейс системы. Для того, чтобы выполнять обмен метаданными между DSpace в соответствии с расширенным профилем, был создан сервис, выполняющий преобразование схем метаданных из внутренней схемы DSpace в схему сервера метаданных и в схему Dublin Core с использованием квалификаторов.

НАУЧНОЕ НАСЛЕДИЕ А.А.ЛЯПУНОВА

- ✘ Для заполнения основного каталога метаданных в соответствии с созданными схемами метаданных используются контролируемые словари из справочного блока системы. В настоящий момент для заполнения отдельных полей были созданы следующие *контролируемые словари*:
 - + персоны – авторитетные файлы;
 - + ключевые термины;
 - + классификационные индексы;
 - + списки учреждений, организаций и предприятий;
 - + типы документов;
 - + географические названия;
 - + типы содержимого документов.