

# Модель электронной библиотеки по научному наследию<sup>1</sup>

## Model of the digital library for scientific heritage

## Модель електронної бібліотеки з науково спадщини

*А. М. Федотов, О. А. Федотова*

*Институт вычислительных технологий СО РАН,*

*ГПНТБ СО РАН,*

*Новосибирск, Россия*

*Anatoly Fedotov and Olga Fedotova*

*Institute of Computational Technologies SB RAS,*

*State Public Scientific and Technical Library of the Russian Academy of Sciences,*

*Novosibirsk, Russia*

*А. М. Федотов, О. А. Федотова*

*Институт обчислювальних технологій СВ РАН,*

*ДПНТБ СВ РАН*

*Новосибірськ, Росія*

Доклад посвящен обсуждению модели электронной библиотеки (ЭБ) по научному наследию и ее архитектуре. Особое внимание уделено функциональным требованиям к ЭБ такого типа, которые определяются, во-первых, информационными потребностями исследователей, а во-вторых, обеспечением надежного и долговременного хранения информации.

**Ключевые слова:** научное наследие, информационная система, электронная библиотека, библиографические базы данных, цифровой депозитарий, научные школы, DSpace, протокол OAI, абстрактный поиск, метаданные, МЕКОФ.

The paper discusses the model digital library (DL) for scientific heritage and its architecture. Special attention is paid to the functional specifications for this type of DL, which are determined, first of all, by researchers information needs, and secondly, by the demand for reliable and long-term preservation of information.

**Keywords:** scientific heritage, information system, electronic library, bibliographic databases, digital repositories, scientific schools, DSpace, OAI, abstract search, metadata, MEKOF.

Доповідь присвячено обговоренню моделі електронної бібліотеки (ЕБ) з наукової спадщини та її архітектури. Особливу увагу приділено функціональним вимогам до ЕБ такого типу, що визначаються, по-перше, інформаційними потребами дослідників, а по-друге, забезпеченню надійного і довготривалого збереження інформації.

**Ключові слова:** наукова спадщина, інформаційна система, електронна бібліотека, бібліографічні бази даних, цифровий депозитарій, наукові школи, DSpace, протокол OAI, абстрактний пошук, метадані, МЕКОФ.

### Введение

Основным требованием, предъявляемым к информационным системам, предназначенных для поддержки научных исследований, является интероперабельность.

Под интероперабельностью любой информационной системы, в том числе и электронной библиотеки, понимается степень ее способности взаимодействовать с другими информационными системами, в том числе и с человеком [1]. Но если при взаимодействии с последним (как с информационной системой) основная нагрузка на обеспечение взаимопонимания ложится на человека, который в состоянии обработать даже очень плохо организованную информацию, то для обеспечения эффективного взаимодействия между информационными системами требуются специальные технологические подходы и общие соглашения. Обеспечение интероперабельности систем

---

<sup>1</sup> Работа выполнена при поддержке РФФИ (проекты 13-07-00258, 12-07-00472, 11-07-00561, 10-07-00302), а также программы «Ведущие научные школы РФ» (грант НШ 6293.2012.9) и проекта ФЦП № «07-514-0022-004»

невозможно без строгого выполнения соответствующих международных стандартов и рекомендаций.

В работе [2] был определен базовый профиль стандартов информационной системы для поддержки научных исследований, организованной в виде электронной библиотеки. Под профилем метаданных понимается адаптация существующей схемы метаданных к нуждам конкретной задачи, решаемой информационной системой [3]. Исходя из анализа существующих форматов метаданных, предназначенных для работы с публикациями, документами и другими информационными ресурсами, можно сделать вывод, что для исследовательской работы с материалами по научному наследию наиболее подходящим является формат ГОСТ 7.19-2001 (МЕКОФ) [4]. По сравнению с другими общеупотребительными форматами метаданных (форматы семейства MARC, см., например, [5,6]) данный формат обладает наиболее полной системой классификации типов документов и других информационных ресурсов и достаточно большим набором словарей-справочников, необходимых для описания и идентификации информационных ресурсов.

В настоящей работе рассматривается технологический подход к созданию типовой модели информационной системы, предназначенной для обеспечения научных исследований. Разрабатываемая модель информационной системы для работы с материалами, имеющих отношение к научному наследию, должна решать задачи долговременного хранения информации, организации абстрактного поиска по атрибутам, организации сбора и обмена метаданными и информацией между удаленными хранилищами информационных ресурсов.

## 1. 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.

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.

В классической монографии [7], изданной ВИНТИ еще в 1976 году, выделено 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» [8].

В интеллектуальных информационных системах в качестве составного компонента выступают рассуждающая система, формализующая правила логического вывода, и интеллектуальный интерфейс. Для уточнения смысла, вкладываемого в термин «факт» применительно к информационным системам по научному наследию, представляется целесообразным использование семиотического подхода, подобно тому, как это было сделано в работе [9] для терминов «информация», «знание», «тезаурус», «онтология». В этой работе, в частности, показано, что данные соответствуют синтаксическому уровню сообщения (в том числе документа), информация (в узком смысле!) – семантическому, а знания – прагматическому. Отсюда вытекает, что функционирование интеллектуальной информационной системы основано на двух противоположных процессах: при пополнении ИС новыми сведениями происходит преобразование семантической информации в данные,

однако непосредственно потребности пользователя удовлетворяет обратный процесс – извлечение из данных нужной пользователю информации и знаний.

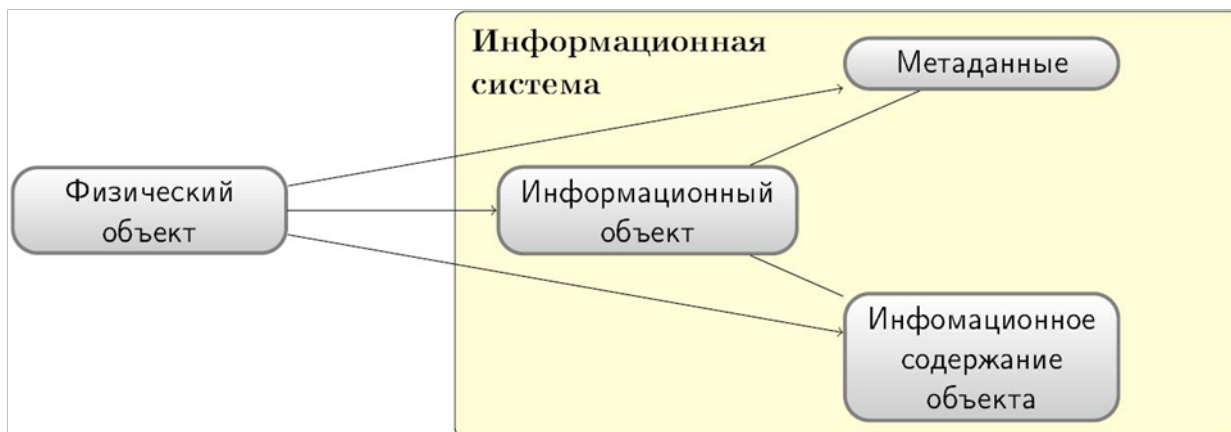


Рис. 1. Структура информационной системы

Ввиду того, что информация в ИС является отображением реальных или материальных сущностей (предметов, процессов, явлений, персон, публикаций, фактов, ключевых терминов 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 (см. рис. 1).

Очевидными являются два необходимых требования, которые можно предъявить к таким информационным системам. Первое требование – необходимость создания и обеспечение системы надежного долговременного хранения цифровых (электронных) документов с сохранением всех смысловых и функциональных характеристик исходных документов. Второе требование – это обеспечение «прозрачного» поиска и доступа к документам пользователей, как для ознакомления, так и для анализа и проведения научной работы [8].

В существующих разработках электронных библиотек, как правило, поиск и доступ к информации обеспечивается только посредством визуальных графических интерфейсов. Это хорошо для пользователя-человека, но очень плохо для пользователя-приложения (например, для проведения различных аналитических исследований<sup>2</sup>).

Для обеспечения функций поиска вне графических интерфейсов требуется поддержка специальных сетевых сервисов и языков запросов. В идеальном случае все информационные системы должны поддерживать единый поисковый профиль и единый язык запросов. Реализация парадигмы абстрактного поиска сегодня существует в виде нескольких моделей организации поисковых сервисов, например, модель Z39.50 [10,11] и более простая модель SRW/SRU [12,13]. Практическая реализация сервисов типа SRW/SRU дает существенно новое качество электронной библиотеки – возможность включения ее ресурсов в глобальные поисковые системы на более высоком уровне, нежели уровень внешней индексации статических WEB страниц другими системами.

## 2. Архитектура информационной системы по научному наследию

Исходя из сформулированных требований, информационная система поддержки исследований по научному наследию должна состоять из системы долговременного хранения и системы управления информацией для организации абстрактного поиска необходимого для анализа и проведения научных работ. Очень важным компонентом технологии работы с научным наследием являются метаданные, содержащие информацию, необходимую для документирования процесса хранения информационных ресурсов. Эти метаданные представляют собой информацию о формате, струк-

<sup>2</sup> Можно делать выборки и вручную, но в наше время это странно.

туре и использовании информационных ресурсов, историю всех операций, в том числе любые изменения, аутентичность, техническую историю, ответственность, права и пр.

Наиболее важным выводом из вышесказанного является то, что информационная модель ЭБ по научному наследию должна быть многоуровневой и состоять как минимум из следующих компонент [14,15]: хранилище данных – репозитарий, сервер метаданных, сервер приложений, словари-справочники (см. рис. 2):

1. *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).

2. *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 [4].

3. *Сервер метаданных* должен обеспечить работу с метаданными – каталогизацию всех информационных ресурсов в соответствии с общепринятыми международными стандартами.

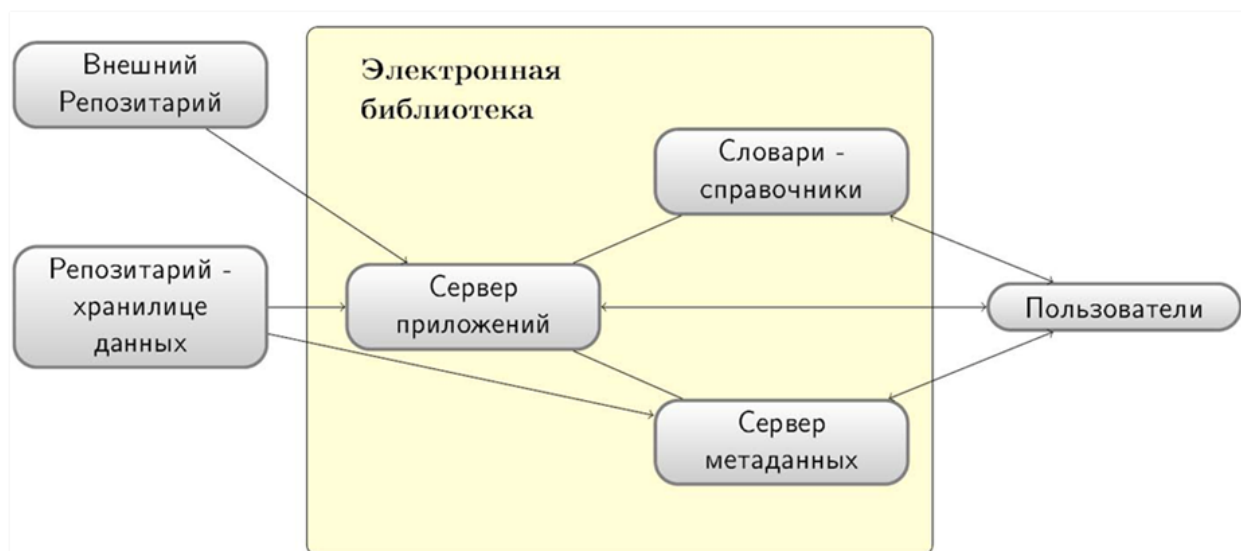


Рис. 2. Архитектура электронной библиотеки

### 3. Реализация

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 ГОСТ 7.19 (МЕКОФ) [4]. Разработанная схема учитывает основные требования этого стандарта и схемы метаданных Dublin Core [16]. Для долговременного хранения документов использовался репозитарий DSpace[17]. 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. Реализован также OAI [18] сервис, который в пакетном

режиме периодически, в соответствии с расписанием, проводит синхронизацию метаданных ЦД и сервера метаданных.

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.

Разработанная модель информационной системы может быть использована как типовая модель системы для работы с документами, связанных с научным наследием, поскольку решает основные задачи, предъявляемые к этим системам: обеспечение системы надежного долговременного хранения цифровых (электронных) документов с сохранением всех смысловых и функциональных характеристик исходных документов; обеспечение «прозрачного» поиска и доступа к документам пользователей как для ознакомления, так и для анализа и проведения научной работы; организация сбора информации по удаленным ЦД, поддерживающих протокол OAI.

## Литература

1. Жижимов О.Л., Федотов А.М. Обеспечение интероперабельности электронных библиотек // Информационные технологии и математическое моделирование в науке, технике и образовании (Бишкек, Кыргызстан, 05.10-09.10.2011): Известия Кыргызского Государственного Технического Университета им.И.Разакова. – №24. Материалы международной конференции. – Бишкек, 2011. – с. 331-335.
2. Федотов А.М., Барахнин В.Б., Жижимов О.Л., Федотова О.А. Технология создания корпоративных информационных систем учета трудов научных работников // Вестник НГУ. Сер.: Информационные технологии. 2011. Т. 9, вып. 2. С. 31-41.
3. ГОСТ Р ИСО / МЭК ТО 10000-2-99. Информационная технология. Основы и таксономия функциональных стандартов. Часть 2. Принципы и таксономия профилей ВОС.
4. ГОСТ 7.19-2001 Система стандартов по информации, библиотечному и издательскому делу. Формат для обмена данными. Содержание записи.
5. MARC Standards / Library of Congress, Network Development and MARC standards Office // The Library of Congress [web-сайт] / Library of Congress. – 2007. <http://www.loc.gov/marc/>.
6. RUSMARC [Электронный ресурс]: Российский коммуникативный формат // Российская Библиотечная Ассоциация [web-сайт] / Российская Библиотечная Ассоциация. <http://www.rba.ru:8101/rusmarc/>
7. Михайлов А.И., Черный А.И., Гиляревский Р.С. Научные коммуникации и информатика. // М.: Наука, 1976. – 436 с.
8. Шокин Ю.И., Федотов А.М., Барахнин В.Б. Проблемы поиска информации. Новосибирск: Наука, 2010. – 198с.
9. Барахнин В.Б., Федотов А.М. Уточнение терминологии, используемой при описании интеллектуальных информационных систем, на основе семиотического подхода // Известия вузов. Проблемы полиграфии и издательского дела. – 2008. № 6. С. 73-81.
10. Жижимов О.Л., Мазов Н.А., Федотов А.М. Некоторые заметки об эволюции цифровых репозитариев традиционных библиотек к полнофункциональным электронным библиотекам // Вестник Владивостокского государственного университета экономики и сервиса. Территория новых возможностей. – №3 (7). – 2010. – с.55-63.
11. ANSI/ISO Z39.50-2003. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification // NISO Press, Bethesda, Maryland, U.S.A. – ISBN 1-880124-55-6. – 267 p.
12. Жижимов О.Л., Мазов Н.А. Принципы построения распределенных информационных систем на основе протокола Z39.50. Новосибирск: ОИГМ СО РАН; ИВТ СО РАН, 2004. 361 с.
13. SRU – Search/Retrieval via URL / The Library of Congress. – USA – <http://www.loc.gov/standards/sru>
14. Федотов А.М. Методологии построения распределенных систем // Вычислительные технологии. 2006. Т. 11. С. 3-17.

15. Жижимов О.Л., Федотов А.М., Федотова О.А. Построение типовой модели информационной системы для работы с документами по научному наследию // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2012. – Т. 10, № 3. С. 5-14.
16. Dublin Core Metadata Initiative [Электронный ресурс] / Dublin Core Metadata Initiative. – 2007. <http://www.dublincore.org/>
17. DSpace [Электронный ресурс]: an open source solution for accessing, managing and preserving scholarly works // dspace.org [web-сайт] / MIT Libraries; HP Labs. – 2007. <http://www.dspace.org/>
18. The Open Archives Initiative Protocol for Metadata Harvesting [Электронный ресурс]: Protocol Version 2.0 of 2002-06-14 // Open Archives Initiative: [web-сайт] / The OAI Executive; OAI Technical Committee. – 2004. <http://www.openarchives.org/OAI/openarchivesprotocol.html>