

УДК 004.912

**А. М. Федотов^{1,2}, А. Т. Байдавлетов², О. Л. Жижимов¹
М. А. Самбетбаева², О. А. Федотова³**

¹ *Институт вычислительных технологий СО РАН
пр. Акад. Лаврентьева, 6, Новосибирск, 630090, Россия*

² *Новосибирский государственный университет
ул. Пирогова, 2, Новосибирск, 630090, Россия*

³ *Государственная публичная научная библиотека СО РАН
ул. Восход, 15, Новосибирск, 630200, Россия*

*fedotov@sbras.ru, baidauletov.at@gmail.com, zhizhim@mail.ru
madina_jgtu@mail.ru, o4f8@mail.ru*

ЦИФРОВОЙ РЕПОЗИТОРИЙ В НАУЧНО-ОБРАЗОВАТЕЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЕ

Статья посвящена сравнительному анализу наиболее популярных систем поддержки цифровых репозиторий и их информационной модели. Обосновывается выбор системы DSpace для хранилища данных научно-образовательной информационной системы.

Ключевые слова: информационная система, электронная библиотека, цифровой репозиторий, хранилище данных, эталонная модель OAIS, протокол OAI-PMH, DSpace, EPrints, Fedora, GreenStone, CDC Invenio.

Введение

В работах [1; 2] были сформулированы основные цели создания информационных систем (ИС), предназначенных для работы научными документами в процессе научно-образовательной деятельности, а именно:

- предоставление научным работникам, преподавателя и студентам быстрого доступа к информационным ресурсам;
- предоставление результатов фундаментальных научных исследований мировому сообществу;
- предотвращение утраты ценных научных коллекций для будущих поколений ученых и студентов;
- создание новых технологий научных исследований и обучения, эффективного инструментария для их проведения.

В информационной системе, направленной на поддержку научно-образовательной деятельности, важно хранить описание всего жизненного цикла информационных ресурсов и иметь возможность восстановить состояние ресурса на любой момент времени. Существуют информационные ресурсы, которые должны быть доступны длительное время. К таким, например, относятся документы и публикации, составляющие научное наследие, документы, имеющие длительную юридическую силу, патенты, мультимедийная информация об исторических событиях, которая может быть востребована через любой период времени. Кроме того, научные отчеты институтов, речи ученых, письма и служебные записки могут

Федотов А. М., Байдавлетов А. Т., Жижимов О. Л., Самбетбаева М. А., Федотова О. А. Цифровой репозиторий в научно-образовательной информационной системе // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2015. Т. 13, вып. 3. С. 68–86.

также иметь огромную историческую значимость, становясь более ценной со временем. Поэтому ИС должна поддерживать возможность длительного и надежного хранения информационных ресурсов с возможностью восстановления их.

Исходя из целей создания ИС предназначенных для поддержки научно-образовательной деятельности и анализа существующих систем, направленных на поддержку научных исследований, можно сформулировать следующие функциональные требования к модели ИС:

- 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;
- 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 ИС должна быть многоуровневой и состоять как минимум из двух компонент [3; 4]: подсистема хранения данных и подсистема сервисов управления информационными ресурсами. Причем функция хранения данных должна быть отделена и не зависеть от других функций и сервисов системы. Подсистема хранения данных – цифровой репозиторий – является одним из важнейших компонентов системы и предназначена только для обеспечения «функции» долговременного хранения информационных ресурсов. Сервисы могут меняться, а данные должны храниться надежно и вечно.

В статье описана информационная модель цифрового («институционального») репозитория, дан сравнительный анализ наиболее популярных систем поддержки институциональных репозитория и обоснован выбор системы DSpace [5] для организации подсистемы долговременного хранения информационных ресурсов в научно-образовательной ИС.

Институциональный репозиторий

Для организации системы долговременного хранения информационных ресурсов международной организацией по стандартизации (ISO) предложен стандарт ISO-14721:2012¹ (OAIS – Open Archive Information System) [6]. Эталонная модель OAIS была использована многими организациями для разработки наборов метаданных и организации крупных хранилищ цифровых объектов.

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

- обеспечение разграниченного доступа к разнородным цифровым объектам (публикациям, изображениям и т. д.);
- организация доступа к информационным ресурсам для мирового сообщества (в том числе с помощью полнотекстового индексирования мировыми поисковыми системами);
- унифицированный доступ к метаданным по стандартным протоколам (поддержка интероперабельности);
- возможность организации единой точки доступа к информационным ресурсам;

¹ http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

- сохранение других информационных ресурсов, в том числе неопубликованных, таких как диссертации, препринты и технические отчеты, программное обеспечение (ПО), мультимедиа и т. д. Институциональный репозиторий – система длительного хранения, накопления и обеспечения надежного доступа в электронной форме интеллектуальной продукции учреждения, в частности, научного или образовательного учреждения.

Институциональные репозитории связаны с вопросами цифровой интероперабельности и с инициативой открытых архивов (ОАИ²), а также частично связаны с понятием электронной библиотеки, т. е. с функциями сбора, хранения, классификации, каталогизации и обеспечения доступа к цифровому контенту. Процесс интеграции цифрового репозитория (хранилища данных) в информационную систему основан на модели агрегирования и распространения метаданных. Применение этой модели закреплено в протоколе OAI Protocol for Metadata Harvesting (далее OAI-PMH) [7], который поддерживается большинством систем, предназначенных для хранения информационных ресурсов.

Информационная модель OAI-PMH

Инициатива «Открытые архивы» (ОАИ) разрабатывает и продвигает стандарты интероперабельности с целью эффективного распространения электронных ресурсов, а также повышения доступности обмена научной информацией. Протокол OAI-PMH имеет простую и гибкую информационную модель (рис. 1).

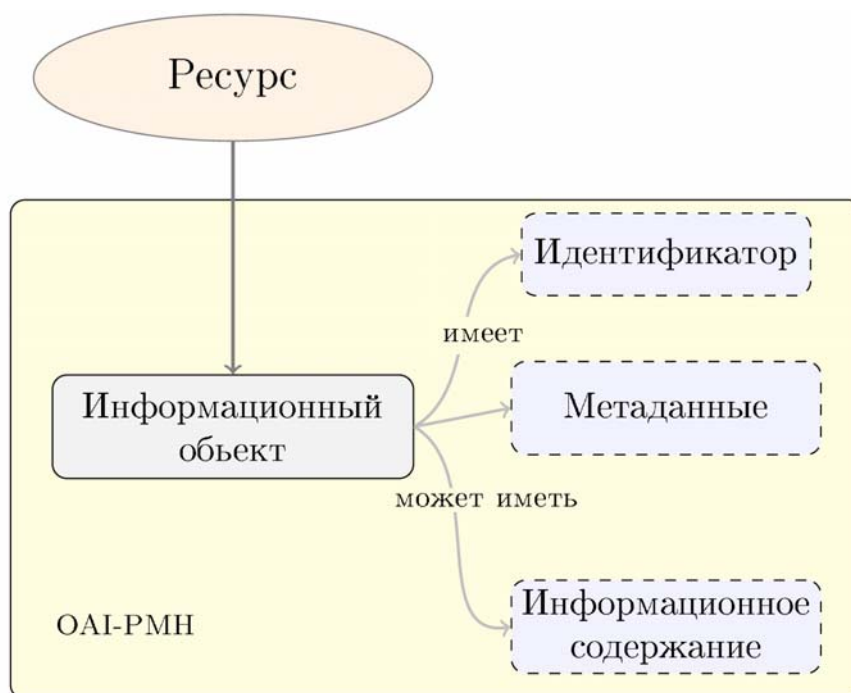


Рис. 1. Информационная модель OAI-PMH:
«Ресурс-Информационный объект-Запись»

Основная сущность, с которой оперирует модель, – это Ресурс (Resource). В общем случае Ресурс интерпретируется в терминах библиографических данных, описывающих ресурс, хотя возможны и другие интерпретации [7; 8]. В информационной системе Ресурсу соответствует Информационный объект (Item – элемент), который является традиционным вторичным информационным объектом, содержащим описание первичного ресурса (в нашем случае публикации или документа).

² OAI – Open Archives Initiative.

Информационный объект концептуально является контейнером, который хранит или динамически генерирует метаданные об отдельном ресурсе в нескольких форматах, каждый из которых может быть собран в виде *записи* OAI-PMH. Каждый информационный объект имеет *идентификатор записи* или уникальный идентификатор. Уникальный идентификатор однозначно определяет Информационный объект в репозитории, он используется в запросах OAI-PMH для извлечения метаданных. Протокол OAI-PMH требует обязательного включения в описание ресурса набора метаданных Дублинское ядро (Dublin Core – DC). Также желательно включать в описание пользовательские наборы метаданных (например, MARC³), расширяющие DC (рис. 2).



Рис. 2. Информационная модель OAI-PMH: «Запись (record)»

Записи описывают ресурс в произвольной схеме метаданных, выраженной на языке XML. Техническая среда OAI-PMH базируется на двух известных требованиях к метаданным: интероперабельность и расширяемость. Требование интероперабельности метаданных решается так, что все провайдеры данных в OAI поставляют метаданные в расширенном формате (например, MODS, MARC, XOAI и т. п.), хотя минимально обязательным является схема Dublin Core [9].

Применение протокола OAI-PMH требует обмена информацией в жестко регламентированных схемах данных. Эти схемы, как правило, не соответствуют схемам, которые реализованы в конкретных информационных системах. Поэтому необходима динамическая конвертация метаданных «на лету» (рис. 3).

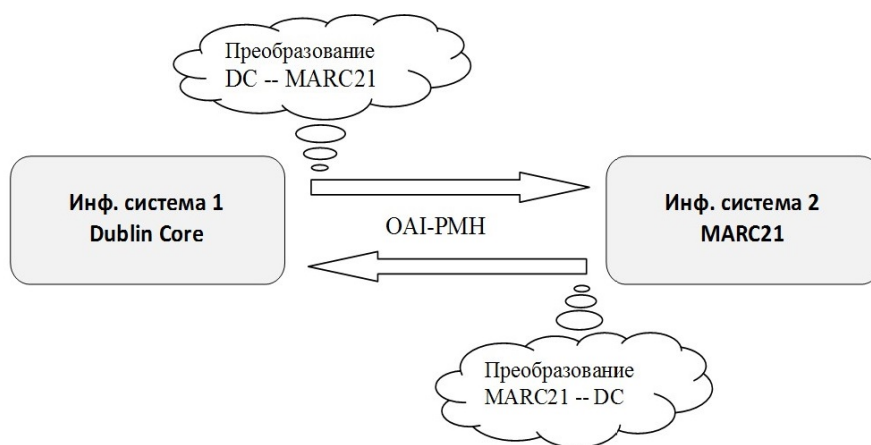


Рис. 3. Схема обмена данными по OAI-PMH

³ Machine-Readable Cataloging (MARC) – <http://www.loc.gov/marc/>.

Уникальный идентификатор указывает на информационный объект (элемент), и все возможные записи, имеющиеся в одном информационном объекте (элементе), совместно используют один и тот же уникальный идентификатор.

Протокол рекомендует использование элемента записи метаданных для связывания записи с идентификатором (URL, URN, DOI, Handle и пр.) ассоциированного документа (объекта) [10]. Для этой цели обязательный формат Дублинское ядро предоставляет атрибут «идентификатор» (DC.Identifier).

Необходимо подчеркнуть, что идентификатор записи не есть идентификатор документа (объекта), поэтому это необходимо для доступа к полному тексту ресурса, описанному записью метаданных.

Технологически OAI-PMH опирается на протокол сбора метаданных, базирующийся на протоколе HTTP. Протокол использует шесть типов запросов и HTTP-методы POST и GET. При этом любой запрос начинается с параметра *verb=*, после которого стоит название команды. Например, запрос *verb=GetRecord*, что означает запрос на получение записи (см. ниже). В общем случае в качестве значения для *verb=* могут быть использованы следующие команды (некоторые имеют дополнительные аргументы):

1. *GetRecord*. Команда используется для поиска индивидуальной записи (метаданных) в репозитории. При этом необходимые дополнительные аргументы – *identifier* – задающий уникальный идентификатор (или ключ требуемой записи), и *metadataPrefix* – формат вывода метаданных в ответ на данный запрос.

```
http://elib.ict.nsc.ru/oai/request?  
verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:elib.ict.nsc  
.ru:ICT/3
```

2. *Identify*. Команда используется для вывода информации о заданном репозитории. В качестве ответа возвращается следующая информация: а) содержательное имя репозитория; б) базовый URL репозитория; в) версия протокола OAI, поддерживаемая репозиторием; г) адрес электронной почты администратора; д) даты создания и обновления. Пример данного запроса:

```
http://elib.ict.nsc.ru/oai/request?verb=Identify
```

3. *ListIdentifiers*. Команда используется для поиска идентификаторов записей, которые представлены в архиве. При помощи этой команды можно выделить те документы, которые были созданы или изменены в последнее время. Например, найти новые поступления в архив. Дополнительные аргументы позволяют искать идентификаторы выборочно, основываясь на их принадлежности к определенной части архива (аргумент *set* позволяет задать коллекции архива) или по временным параметрам (аргументы *from* и *until* ограничивают выборку документов по временному промежутку, может быть указан только один из них, или ни одного). Существует также аргумент *resumptionToken*. Если возвращаемых на запрос данных слишком много, то этот параметр позволяет регулировать показ следующей страницы результатов запроса. Пример запроса:

```
http://elib.ict.nsc.ru/oai/request?verb=ListIdentifiers&metadata  
Prefix=oai_dc
```

4. *ListMetadataFormats*. Команда используется для получения списка форматов метаданных, в которых могут быть представлены выходные данных. Пример запроса:

```
http://elib.ict.nsc.ru/oai/request?verb=ListMetadataFormats  
<ListMetadataFormats>  
<metadataFormat><metadataPrefix>rusmarc</metadataPrefix>
```

```
<schema>http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd</schema><metadataNamespace>http://www.loc.gov/MARC21/slim</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>mods</metadataPrefix>  
<schema>http://www.loc.gov/standards/mods/v3/mods-3-1.xsd</schema><metadataNamespace>http://www.loc.gov/mods/v3</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>ore</metadataPrefix>  
<schema>http://tweety.lanl.gov/public/schemas/2008-06/atom-tron.sch</schema><metadataNamespace>http://www.w3.org/2005/Atom</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>mets</metadataPrefix>  
<schema>http://www.loc.gov/standards/mets/mets.xsd</schema>  
<metadataNamespace>http://www.loc.gov/METS/</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>mekof</metadataPrefix>  
<schema>http://elib.sbras.ru/MEKOF/slim/MEKOFslim.xsd</schema>  
<metadataNamespace>http://elib.sbras.ru/MEKOF/slim</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>xoai</metadataPrefix>  
<schema>http://www.lyncode.com/schemas/xoai.xsd</schema>  
<metadataNamespace>http://www.lyncode.com/xoai</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>dim</metadataPrefix>  
<schema>http://www.dspace.org/schema/dim.xsd</schema>  
<metadataNamespace>http://www.dspace.org/xmlns/dspace/dim</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>uketd_dc</metadataPrefix>  
<schema>http://naca.central.cranfield.ac.uk/ethos-oai/2.0/uketd_dc.xsd</schema>  
<metadataNamespace>http://naca.central.cranfield.ac.uk/ethos-oai/2.0/</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>qdc</metadataPrefix>  
<schema>http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd</schema><metadataNamespace>http://purl.org/dc/terms/</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>oai_dc</metadataPrefix>  
<schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>  
<metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>rdf</metadataPrefix>  
<schema>http://www.openarchives.org/OAI/2.0/rdf.xsd</schema>  
<metadataNamespace>http://www.openarchives.org/OAI/2.0/rdf/</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>marc</metadataPrefix>
```

```
<schema>http://www.loc.gov/standards/marcxml/schema/MARC21slim.xsd</schema><metadataNamespace>http://www.loc.gov/MARC21/slim</metadataNamespace></metadataFormat>
```

```
<metadataFormat><metadataPrefix>etdms</metadataPrefix>
<schema>http://www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd</schema><metadataNamespace>http://www.ndltd.org/standards/metadata/etdms/1.0/</metadataNamespace></metadataFormat>
</ListMetadataFormats>
```

5. *ListRecords*. Команда используется для получения записей, входящих в архив. Дополнительные аргументы позволяют выполнять выборочное получение записей, основываясь на принадлежности записи к конкретной коллекции в репозитории, или по временным параметрам. Перечень аргументов и их значение – такие же, как у команды *ListIdentifiers* и включает *from*, *until*, *set* и *resumptionToken*. Пример запроса:

```
http://elib.ict.nsc.ru/oai/request?
verb=ListRecords&metadataPrefix=oai_dc
```

Фрагмент ответа:

```
<ListRecords>
<record><header><identifier>oai:elib.ict.nsc.ru:ICT/3</identifier>
<datestamp>2015-04-18T11:04:23Z</datestamp>
<setSpec>com_ICT_1</setSpec>
<setSpec>col_ICT_2</setSpec></header>

<metadata><oai_dc:dc
xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
xmlns:doc="http://www.lyncode.com/xoai"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/"  xsi:schemaLocation= "
http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">

  <dc:title>Некоторые заметки об эволюции цифровых репозиториях
традиционных библиотек к полнофункциональным электронным
библиотекам</dc:title>
  <dc:creator>Жижимов, Олег Львович</dc:creator>
  <dc:creator>Мазов, Николай Алексеевич</dc:creator>
  <dc:creator>Федотов, Анатолий Михайлович</dc:creator>
  <dc:date>2014-02-25T05:32:34Z</dc:date>
  <dc:date>2015-04-01T03:28:31Z</dc:date>
  <dc:date>2014-02-25T05:32:34Z</dc:date>
  <dc:date>2015-04-01T03:28:31Z</dc:date>
  <dc:date>2010</dc:date>
  <dc:type>Article</dc:type>
  <dc:identifier>Жижимов О.Л., Мазов Н.А., Федотов А.М. Некоторые
заметки об эволюции цифровых репозиториях традиционных библиотек к
полнофункциональным электронным библиотекам // Вестник
Владивостокского государственного университета экономики и сервиса.
Территория новых возможностей. – 2010. – Т.3. – № 7. – С.55-63. –
ISSN 2073-3984.</dc:identifier>
  <dc:identifier>2073-3984</dc:identifier>
```

```
<dc:identifier>http://elib.ict.nsc.ru/jspui/handle/ICT/3</dc:identifier>
<dc:language>ru</dc:language>
<dc:relation>Вестник Владивостокского государственного
университета экономики и сервиса. Территория новых
возможностей</dc:relation>
</oai_dc:dc></metadata></record>
</ListRecords>
```

6. *ListSets*. Команда используется для получения структуры коллекций репозитория. Например, при запросе

```
http://elib.ict.nsc.ru/oai/request?verb=ListSets
```

получим в ответ список всех доступных разделов и коллекций репозитория:

```
<ListSets>
<set><setSpec>com_ICT_1</setSpec>

<setName>01 - Труды сотрудников</setName></set>
<set><setSpec>com_ICT_273</setSpec>

<setName>02 - Техническая и нормативная
документация</setName></set>
<set><setSpec>com_ICT_274</setSpec>

<setName>02.01 - Стандарты в информационных
технологиях</setName>
</set><set><setSpec>com_ICT_331</setSpec>

<setName>03 - Литература по информатике</setName></set>
<set><setSpec>com_ICT_754</setSpec>

и так далее...

</ListSets>
```

Системы поддержки цифровых хранилищ

В настоящий момент в мире насчитывается более десятка систем поддержки цифровых хранилищ (институциональных репозиториев). Согласно данным сайта OpenDOAR⁴ (The Directory of Open Access Repositories) большинство институциональных репозиториев основаны на эталонной модели RM OAIS, созданы на свободном программном обеспечении и поддерживают протокол сбора метаданных OAI-PMH [11; 12].

Наиболее популярные из них, это DSpace⁵ (44% установок), Eprints⁶ (14% установок), Fedora⁷, Greenstone⁸, CDC Invenio⁹ (рис. 4).

Наиболее популярное программное обеспечение для открытых цифровых репозиториев по России (рис. 5) – это DSpace (71% установок), EPrints (8% установок).

⁴ <http://www.openoar.org/>

⁵ <http://www.dspace.org/>

⁶ <http://www.eprints.org/>

⁷ <http://www.fedoracommons.org/>

⁸ <http://www.greenstone.org/>

⁹ <http://invenio-software.org/>

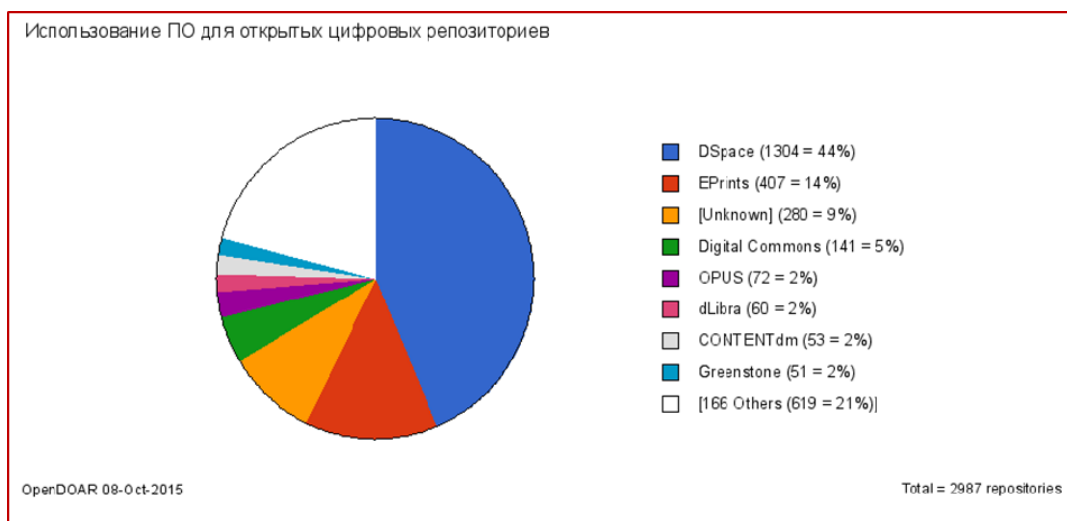


Рис. 4. Использование ПО для открытых цифровых репозиториях

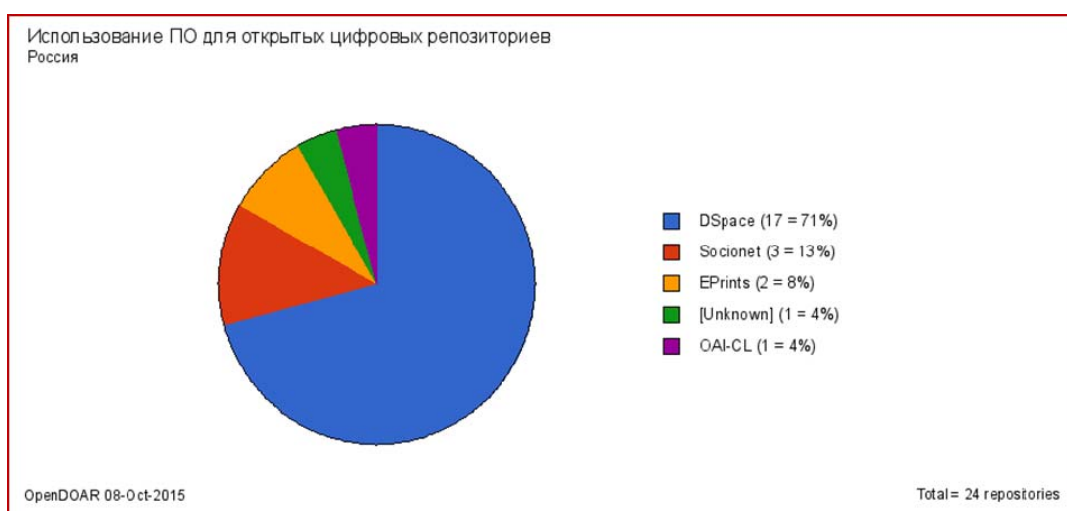


Рис. 5. Использование ПО для открытых цифровых репозиториях в России

DSpace – это самое популярное в академической среде программное обеспечение для создания архива электронных ресурсов (цифрового репозитория). *DSpace* обеспечивает платформу для долгосрочного хранения цифровых материалов, используемых в академических исследованиях. Платформа *DSpace* разрабатывалась совместно компанией Hewlett-Packard и библиотеками MIT (Massachusetts Institute of Technology). В ноябре 2002 года система была запущена как действующая служба, поддерживаемая библиотеками MIT. Движение Scholarly Communication (Общение в мире науки) оказало влияние на развитие *DSpace*, вследствие чего конфигурация по умолчанию направлена на поддержку научных публикаций [5; 14].

DSpace – открытое (лицензия BSD) кроссплатформенное Java приложение. Для хранения метаданных используется СУБД Oracle или PostgreSQL. Для базовой организации данных зафиксирована определенная модель данных, основанная на схеме Dublin Core¹⁰ и ее расширениях, система хранит (конвертирует) и индексирует метаданные в разнообразных

¹⁰ <http://dublincore.org/>

форматах (DIM¹¹, MODS¹², METS¹³, QDC¹⁴, MARC¹⁵, ORE¹⁶ и др., список форматов может быть расширен добавлением новых, в том числе собственной генерации, конвертеров). Система хранит информацию о пользователях системы и поддерживает авторизацию и разграничивает доступ к содержимому репозитория. Кроме того, такие функции как депонирование и редакторская проверка привязаны к пользователям.

DSpace работает со стандартными для библиотечной сферы протоколами OAI-PMH, OpenURL[13] и SWORD [15; 16]. С 2009 г. DSpace поддерживается сообществом DURASpace, которое образовано путем слияния двух проектов цифровых репозиториях DSpace Foundation и Fedora Commons.

Fedora (Flexible Extensible Digital Object Repository Architecture) – разработан исследователями из Корнельского университета в качестве платформы для хранения, управления и доступа к цифровому контенту (цифровым объектам) [17].

Fedora определяет набор абстракций для выражения цифровых объектов, отношения между цифровыми объектами, связи и поведение цифровых объектов. В отличие от DSpace, Fedora больше подходит для хранения произвольных цифровых объектов, например программного обеспечения.

Ядро репозитория Fedora предоставляет набор веб-сервисов с четко определенными API. Кроме того, Fedora предоставляет широкий спектр вспомогательных сервисов и приложений, включая поиск, поддержку OAI-PMH, обмен сообщениями, управление клиентами и многое другое.

Метаданные о содержании могут быть представлены в формате METS, FOXML [18]. Обеспечивается поддержка RDF.

EPrints – это программное обеспечение открытых ресурсов для построения хранилищ цифровых ресурсов, первоначально предназначенный для научных публикаций. EPrints был разработан при университете Саузхэмптона.

EPrints – вторая в академическом мире по популярности после DSpace система, которая используется для формирования и управления открытыми архивами и предназначена для создания архивов научных исследований с большим разнообразием информационных ресурсов (научные статьи, отчеты, диссертации, монографии, учебно-методические пособия, материалы конференций, данные результатов экспериментов и наблюдений и т. п.) [19].

Метаданные о содержании могут быть представлены в формате METS, DC, MODS, BibTeX¹⁷, EndNote¹⁸.

Открытые архивы, созданные в среде EPrints, поддерживают протокол обмена метаданными OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting) и SWORD, Atom¹⁹, который обеспечивает глобальные услуги доступа и поиска.

*Greenstone*²⁰ – свободно распространяемое программное обеспечение для создания и поддержания институциональных репозиториях открытого доступа (цифровых онлайн библиотек). Оно является открытым и многоязычным, выпускается под лицензией GNU.

Greenstone – разработано в рамках Проекта новозеландской цифровой библиотеки при Университете Вайкато в сотрудничестве с ЮНЕСКО и Неправительственной организацией гуманитарной информации.

Основная схема данных Dublin Core с квалификаторами, основные форматы документов HTML и MS Word. Для программного доступа к ресурсам имеется собственный API, отличный от стандартных протоколов доступа к цифровым репозиториям таких, как OAI или

¹¹ DSpace Intermediate Metadata (DIM) – внутренний формат данных DSpace.

¹² Metadata Object Description Schema (MODS) – <http://www.loc.gov/standards/mods/>

¹³ Metadata Encoding and Transmission Standard (METS) – <http://www.loc.gov/METS>

¹⁴ Qualified Dublin Core (QDC).

¹⁵ Machine-Readable Cataloging (MARC) – <http://www.loc.gov/marc/>

¹⁶ Object Reuse and Exchange (ORE) – <http://www.openarchives.org/ore/1.0/toc/>

¹⁷ <http://www.bibtex.org/>

¹⁸ <http://endnote.com/>

¹⁹ Atom – общее название двух связанных веб-технологий: формата для описания ресурсов на веб-сайтах и протокола для их публикации.

²⁰ <http://www.greenstone.org/>

Сравнительная характеристика ПО цифровых депозитариев

Характеристика	DSpace	EPrints	Fedora commons	Greenstone	CDS Invenio
Год создания	2002	2000	1997	1997	1993
Организация-разработчик	Штат. Массачусетс, США (Массачусетский технологический институт / HP), с 2009 года Duraspace	Университет Саутгемптона, Великобритания	Корнельский университет, Университет Вирджинии, США, С 2009 года Duraspace	Университет Вайкато, Новая Зеландия	ЦЕРН / CERN, Швейцария
Лицензия	BSD (open source license)	GNU (General Public License)	Apache License, Version 2.0	GNU (General Public License)	GNU (General Public License)
Необходимое программное окружение	<ul style="list-style-type: none"> • Java • Apache Tomcat • PostgreSQL • Unix / Linux/Windows 	<ul style="list-style-type: none"> • Perl • Apache • MySQL • Unix / Linux 	<ul style="list-style-type: none"> • Java • Apache Tomcat • PostgreSQL • (MySQL) • Unix / Linux 	<ul style="list-style-type: none"> • PERL, GDBM, MG • Unix / Linux 	<ul style="list-style-type: none"> • Python (PHP, Common LISP) • Apache • MySQL • Unix / Linux/Windows
Способы организации цифрового архива	Организация различных иерархических коллекции цифровых объектов. Структура и уровень вложенности разделов и коллекций может быть любой. Цифровой объект может быть зарегистрирован в нескольких коллекциях одновременно.	Все записи эквивалентны и являются одноуровневыми (не составляют иерархию)	Состоит из направленного ациклического графа ресурсов, где ветви представляют связь «родитель-потомок»	Иерархическая структура в виде вложенных разделов (sections) (разделы, подразделы, подподразделы и т.д.) отображает содержательную структуру документа. Каждый из разделов, в свою очередь, состоит из одного или нескольких абзацев (paragraphs)	Дерево коллекции строится на основе классификации наполнения и имеет иерархическую структуру
Схема метаданных	Метаданные в схеме Dublin Core	С каждым типом документа связан	Метаданные в схеме Dublin Core	Метаданные в схеме Dublin Core	Метаданные в схеме MARC21, XML.

Продолжение таблицы

Характеристика	DSpace	EPrints	Fedora commons	Greenstone	CDS Invenio
		отдельный набор полей метаданных, возможность динамически генерировать метаданные в различных форматах из внутреннего представления			Структура метаданных в любой момент может быть расширена и адаптирована
Пользователи	Группы пользователей: аноним (просмотр и поиск документов), депоненты, администраторы, участники процесса депонирования, подписчики и пользователи с разрешением просматривать не общедоступные элементы.	Четыре группы пользователей: минимальный пользователь, депонент, редактор, администратор	Группы авторизованных пользователей для доступа цифровых объектов	Пользователь в Greenstone принадлежит к одной из двух predeterminedных пользовательских групп: администратор или разработчик коллекции. Первая группа пользователей имеет право создавать и удалять пользователей, в то время как второй создает и обновляет коллекции. Конечные пользователи имеют доступ ко всем коллекциям и документам	Группы пользователей определяются их правами доступа
Поиск и просмотр	Просмотр и поиск документов в системе, могут выполняться анонимно, но чтобы выполнить внесение документов пользователю нужно	Индексирует все имеющиеся файлы во всех архивах, что позволяет реализовать базовый (simple search) и расширенный (advanced search) поиск	По умолчанию индексация предусмотрена для набора элементов метаданных Dublin core. что позволяет реализовать базовый	По умолчанию пользовательский веб-интерфейс обеспечивает просмотр и поиск в коллекции, перейдя в иерархические объекты. Для просмотра	- Простой поиск; - Расширенный многокритериальный поиск с сортировкой по этим критериям; - Возможность ввести указания для поиска;

Характеристика	DSpace	EPrints	Fedora commons	Greenstone	CDS Invenio
	зарегистрироваться. Поисковые индексы, построенные в соответствии с заданными правилами на основе полей метаданных и контента цифровых объектов. Для хранения и индексирования цифровых объектов DSpace использует Apache SOLR.	Базовый поиск дает возможность выполнять запрос по основным группам метаданных. Расширенный поиск в отличие от базового дает возможность выполнить запрос по отдельным данным.	встроенный поиск и расширенные поиски GSearch, PrOAI. Механизма просмотра не предусмотрено.	коллекции используются определенный перечень метаданных: перечень авторов, названий, дат, иерархические классификационные структуры и т.д.	- Навигация по репозиторию возможна по типу или тематике документа;
Интероперабельность	Поддерживает OAI-PMH, протокол OpenURL и SWORD.	Поддерживает OAI-PMH и SWORD, Atom	Поддерживает OAI-PMH и SWORD	Поддерживает OAI-PMH и протокол Z39.50.	Поддерживает OAI-PMH
Обменные форматы	DIM, MODS, METS, QDC, MARC, ORE, XOAI	METS, MPEG-21, DC, MODS, BibTeX, EndNote	METS, FOXML, The Atom	Qualified DC, RDF, MARC, BibTeX	MARC21
Аутентификация	Аутентификация обеспечивается через пользовательские пароли, IP-адреса, сертификаты X.509, протоколы LDAP или Shibboleth*	Аутентификация обеспечивается через пользовательские пароли, IP-адреса, протоколы LDAP или Shibboleth.	Аутентификация обеспечивается через LDAP или пользовательские пароли	Аутентификация обеспечивается через пользовательские пароли	Механизм управления доступом на основе ролей (RBAC), где пользователи принадлежат нескольким группам согласно их роли в системе.
Сохранность данных	<ul style="list-style-type: none"> Резервное копирование контента. Программа LOCKSS-compliant** 	Резервного копирования экспортируются в файлы формата XML	<ul style="list-style-type: none"> Резервное копирование контента. Программа LOCKSS-compliant 	Резервное копирование осуществляется администратором системы	Резервного копирования экспортируются в файлы формата XML

* Shibboleth разработан в рамках проекта Internet2 и используются университетами США для меж-университетской аутентификации и авторизации и единого доступа к ресурсам, распределенными между университетами.

** LOCKSS (Lots of Copies Keeps Stuff Safe) – метод сохранения информации; опирается на обилие сохраненных копий на многих серверах. Программа LOCKSS, основанная на разработке библиотеки Стэнфордского университета, предоставляет библиотекам и издателям прекрасное дешевое средство сохранения цифровой информации и предоставления доступа к ней; разработка реализована на открытом программном коде.

SRU/SRW, однако есть поддержка протокола Z39.50. Для организации просмотра материала предусмотрено использование внутренних классификаторов [19, 20].

*CDS Invenio*²¹ (ЦЕРН / CERN²², Швейцария) – интегрированная электронная библиотечная система.

Представляет собой набор приложений для построения и управления автономным сервером электронной библиотеки. Программное обеспечение бесплатное, распространяемое под лицензией GNU General Public License. Технология, предлагаемая данным продуктом, покрывает все аспекты поддержки электронной библиотеки, совместима по протоколу OAI-PMH, использует формат MARC21 как основной библиографический стандарт. Система CDS Invenio является комплексным решением управления репозиториями документов средних и больших объемов.

Посредством CDS Invenio создан и поддерживается архив публикаций сервера документов CERN (CERN Document Server). В CERN CDS Invenio управляет более чем 500 коллекциями данных, состоящих из более чем 800 000 библиографических записей и 350 000 полнотекстовых документов, покрывая препринты, статьи, книги, журналы, фотографии, видеоматериалы и др. Помимо CERN, CDS Invenio в настоящее время инсталлирована и используется в 14-ти научных и образовательных учреждениях мира [21].

CDS Invenio поддерживает протокол сбора мета данных Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).

В таблице (см. выше) рассматривается сравнительная характеристика выше описанных 5 систем: DSpace, EPrints, Fedora, Greenstone, CDC Invenio.

Заключение

В качестве репозитория для ИС была выбрана система DSpace. Выбор обусловлен тем, что она является самой популярной в мире и уже эксплуатируется в СО РАН (на рис. 6 показана схема использования репозитория), а также в ряде других институтов и университетов России на протяжении десяти лет. Кроме того, система DSpace имеет ряд привлекательных особенностей:

1. Организовывать хранение цифровых объектов (изображения, медиафайлы, документы в различных форматах и пр.). Версия 5.0 DSpace изначально «понимает» более 70 типов файлов, но это количество может быть легко расширено в зависимости от потребностей конкретного репозитория.

2. Снабжать эти объекты метаданными в соответствии с различными схемами данных. Для базовой организации метаданных зафиксирована определенная модель, основанная на схеме Dublin Core (<http://dublincore.org/documents/dcmi-terms/>) и ее расширениях (<http://purl.org/dc/terms/>). Возможно использование других дополнительных схем, в том числе определенных локально. Для хранения метаданных DSpace использует реляционную СУБД PostgreSQL или Oracle.

3. Создавать поисковые индексы, построенные в соответствии с заданными правилами на основе полей метаданных и контента цифровых объектов, например, текста, извлеченного из PDF документов. Для хранения и индексации цифровых объектов DSpace использует Apache SOLR.

4. Организовывать различные иерархические коллекции цифровых объектов. Структура и уровень вложенности разделов и коллекций может быть любой. Цифровой объект может быть зарегистрирован в нескольких коллекциях одновременно.

5. Использовать WEB-интерфейсы для навигации по иерархическим коллекциям, поиска, просмотра метаданных и цифровых объектов, для управления и исполнения административных функций. DSpace предоставляет две группы интерфейсов: на основе JSP и XML.

6. Контролировать доступ к содержимому репозитория и к его функциям (депонирование, редакторская проверка и т. п.) с различной степенью детализации.

²¹ <http://invenio-software.org/>

²² <http://cds.cern.ch/>

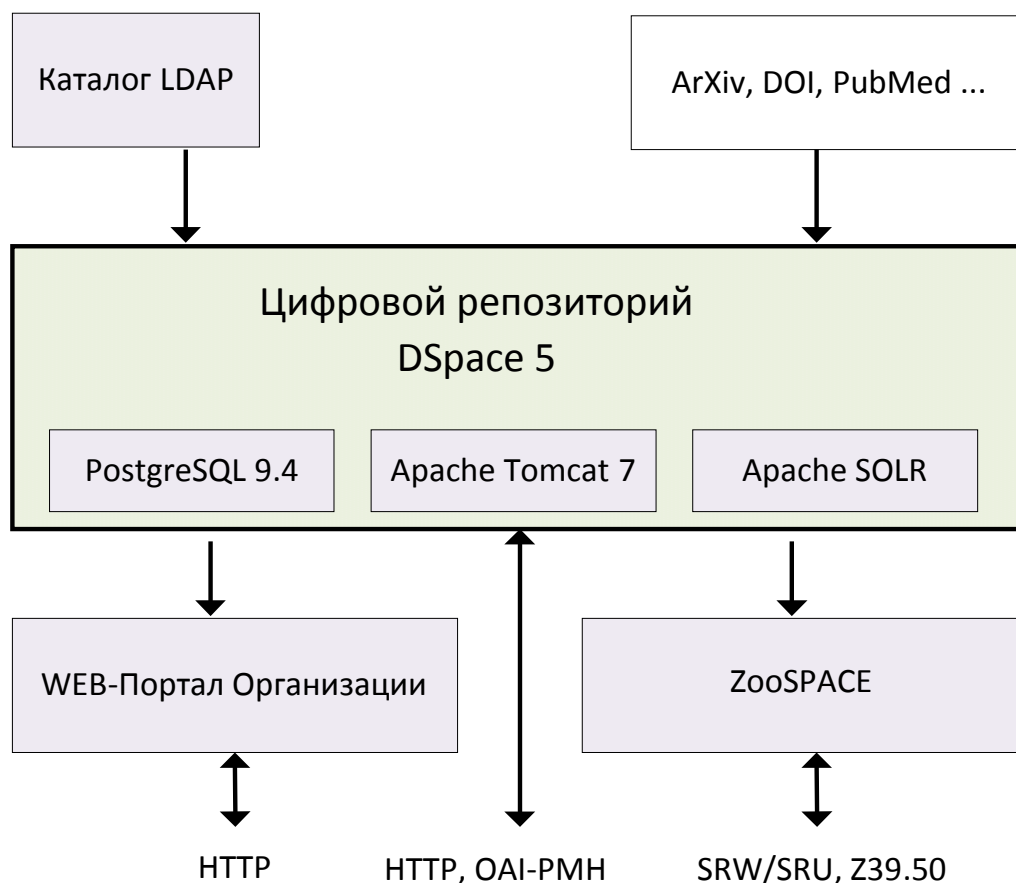


Рис. 6. Схема использования репозитория в ИБТ СО РАН

7. Поддерживать список пользователей репозитория и групп пользователей.

8. Выполнять аутентификацию пользователей, используя различные технологии (локальная, на основе LDAP, и др.). Система хранит информацию о пользователях, поддерживает авторизацию и разграничивает доступ к содержимому репозитория по группам, сетевым адресам и на основе протокола LDAP, 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.

9. Импортировать и экспортировать метаданные и цифровые объекты в соответствии с общепринятыми протоколами и форматами. Наиболее часто используются XML-ориентированные протоколы OAI-PMH, OAI-ORE. При этом сервер, поддерживающий эти протоколы, включен в базовую конфигурацию DSpace. Используя перечисленные протоколы, можно организовывать синхронизацию данных между различными репозиториями. Для базовой организации данных зафиксирована определенная модель данных (DIM – внутренний формат данных DSpace), основанная на схеме Dublin Core и ее расширениях. При некотором напряжении при помощи этой схемы можно отобразить основные элементы всех используемых в настоящий момент схем данных. Система с помощью фильтров метаданных, которые используются для преобразования метаданных из внутренней схемы в схемы, пригодные для обмена метаданными с внешними системами на основе XSLT²³ трансформаций, позво-

²³ <http://www.w3.org/TR/xslt/>

ляет конвертировать и индексировать метаданные в разнообразных форматах (MODS, METS, QDC, MARC и др.). Список форматов может быть расширен добавлением новых, в том числе собственной генерации, конвертеров, например, для формата МЕКОФ²⁴.

10. Организовывать пакетный ввод-вывод данных.

11. Организовывать каталогизацию новых объектов методом заимствования из других репозиторий и баз данных. Список доступных внешних источников может быть расширен изменением конфигурации и / или добавлением новых модулей.

12. Контролировать вводимые при каталогизации данные в избранные поля, используя внутренние и внешние справочники. Наибольший интерес представляет возможность работы с authority данными Библиотеки Конгресса США и ORCID при помощи соответствующих модулей. Каждому типу источника соответствует свой модуль. Настройка существующих модулей осуществляется через конфигурационный файл. Список модулей для авторитетного контроля может быть расширен созданием дополнительных модулей.

13. Мигрировать на новые версии без нарушения общей структуры репозитория.

Для более полного соответствия локальным требованиям в базовую систему DSpace внесены многочисленные изменения (расширение схем данных, расширение номенклатуры обменных форматов, возможность работы с географической информацией, авторитетный контроль и пр.)

В используемой модернизированной системе доступ к данным репозитория возможен не только через WEB-интерфейсы DSpace, но и по протоколам OAI-PMH, SOLR, SRW/SRU, Z39.50. При этом поддержка SRW/SRU и Z39.50 обеспечивается связью DSpace с системой ZooSPACE [22; 23] (см. рис. 6).

Поддержка перечисленных протоколов позволяет разрабатывать собственные информационные системы, взаимодействующие с хранилищем данных, построенном на основе DSpace. В качестве примера на рис. 7 показана страница журнала «Вестник НГУ. Серия: Информационные технологии», полученная из DSpace.

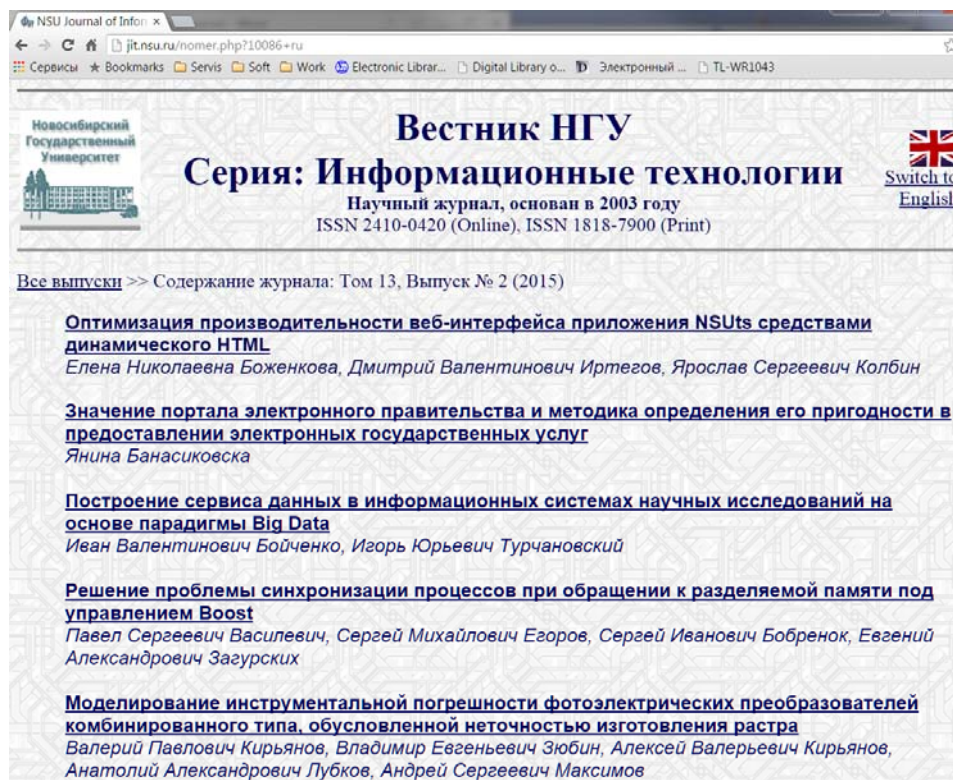


Рис. 7. Образец выдачи информации из DSpace

²⁴ Этот формат наиболее привлекателен в России, так как подержан ГОСТом.

Список литературы

1. Федотов А. М., Федотова О. А. Модель информационной системы для поддержки научно-образовательной деятельности // Вычислительные и информационные технологии в науке, технике и образовании ВИТ-2013: Материалы Междунар. конф. 2013. Т. 2: Вычислительные технологии: Восточно-Казахстанский государственный технический университет. С. 249–265.
2. Федотов А. М., Барахнин В. Б., Жижимов О. Л., Федотова О. А. Модель информационной системы для поддержки научно-педагогической деятельности // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2014. Т. 12, № 1. С. 89–101.
3. Федотов А. М. Методологии построения распределенных систем // Вычислительные технологии. 2006. Т. 11. С. 3–17.
4. 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, № 2. С. 5–14.
5. DSpace: an open source solution for accessing, managing and preserving scholarly works. [Электронный ресурс] / MIT Libraries; HP Labs. – 2007. <http://www.dspace.org/>
6. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, Issue 2, CCSDS 650.0-M-2 (Magenta Book) June 2012. Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
7. 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/>).
8. Резниченко В. А., Проскудина Г. Ю., Кудим К. А. Концептуальная модель электронной библиотеки // Тр. XI Всерос. науч. конф. «RCDL'2009», Россия, г. Петрозаводск (Карелия) 17–21 сентября 2009 г. Петрозаводск, 2009. С. 23–31.
9. Lagoze C., Van de Sompel H. The open archives initiative: building a lowbarrier interoperability framework // Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. New York, NY, USA, 2001. P. 54–62.
10. Sun S., Lammom L., Boesch B. Handle System Overview (RFC 3650). Corporation for National Research Initiatives. 2003. Retrieved from <http://www.ietf.org/rfc/rfc3650.txt>
11. Directory of Open Access Repositories – OpenDOAR. (n.d.). Retrieved August 10, 2013. From <http://opendoar.org/>
12. Goh D. H.-L., Chua A., Khoo D. A., Khoo E. B.-H., Mak E. B.-T., Ng M. W.-M. A checklist for evaluating open source digital library software. Online Information Review, 2006. № 30 (4), P. 360–379.
13. ANSI/NISO Z39.88-2004 (R2010) – The OpenURL Framework for Context-Sensitive Services. National Information Standards Organization. 2010 [2005]. ISBN 978-1-880124-61-1.
14. Кудим К. А., Проскудина Г. Ю., Резниченко В. А. Создание научных электронных библиотек с помощью системы DSpace // Проблемы программирования. 2007. № 3. С. 49–60.
15. Lewis S., de Castro P., Jones R. SWORD: Facilitating Deposit Scenarios. D-Lib Magazine, 2012. № 18 (1/2).
16. SWORD V2 Specifications. (n.d.). Retrieved August 10, 2013. from <http://swordapp.org/sword-v2/sword-v2-specifications/>
17. Pyrounakis G., Saidis K., Nikolaidou M., Karakoidas V. Introducing Pergamos// A Fedora-Based DL System Utilizing Digital Object Prototypes. In Research and Advanced Technology for Digital Libraries. Berlin; Heidelberg: Springer, 2006. P. 500–503.
18. Library of Congress. (n.d.). METS: An Overview & Tutorial. Retrieved from <http://www.loc.gov/standards/mets/METSOverview.v2.html>
19. Tramboo S., Humma H., M Shaft S., Gul S. A study on the Open Source Digital Library Software: Special Reference to DSpace, EPrints and Greenstone // International Journal of Computer Applications. 2012. № 59 (16). P. 1–9.
20. Резниченко В. А., Проскудина Г. Ю., Овдий О. М. Создание цифровой библиотеки коллекций периодических изданий на основе Greenstone. Электронные библиотеки. 2005. Вып. 6. <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>.

21. CERN Document Server. (n.d.). Retrieved August 10, 2013. from <http://cds.cern.ch/>
22. Жижимов О. Л., Федотов А. М., Шокин Ю. И. Платформа ZooSPACE – организация доступа к разнородным распределенным ресурсам // Электронные библиотеки. 2014. Т. 17. № 2.
23. Жижимов О. Л., Федотов А. М., Шокин Ю. И. Технологическая платформа массовой интеграции гетерогенных данных // Вестн. Новосиб. гос. ун-та. Серия: Информационные технологии. 2013. Т. 11, № 1. С. 24–41.

Материал поступил в редколлегию 05.06.2015

**A. M. Fedotov, A. T. Baidavletov, O. L. Zhizhimov
M. A. Sambetbayeva, O. A. Fedotova**

*Institute of Computational Technologies SB RAS
6 Acad. Lavrentjev Ave., Novosibirsk, 630090, Russian Federation*

*Novosibirsk State University
2 Pirogov Str., Novosibirsk, 630090, Russian Federation*

*State Public Scientific Technological Library SB RAS
15 Voskhod Str., Novosibirsk, 630200, Russian Federation*

*fedotov@sbras.ru, baidavletov.at@gmail.com, zhizhim@mail.ru
madina_jgtu@mail.ru, o4f8@mail.ru*

DIGITAL REPOSITORY OF SCIENTIFIC AND EDUCATIONAL INFORMATION SYSTEM

The article is devoted to the comparative analysis of the most popular support systems of digital repositories and their information model. The DSpace system choice for the data warehouse of a scientific and educational information system is justified.

Keywords: information system, electronic library, digital repository, Data Warehouse, the reference model OAIS, the Protocol OAI-PMH, DSpace, EPrints, Fedora, GreenStone, CDC Invenio.

References

1. Fedotov A.M., Fedotova O.A. Model of the information system to support research and educational activities // Computational and informational technologies in science, engineering and education – 2013 (VIT-2013) proceedings of the International conference – 2013. – T.2. – T.: Computational Technologies: East Kazakhstan state technical University. – C.249-265. (in Russian)
2. 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. pp. 89–101. – ISSN 1818-7900. (in Russian)
3. Fedotov A.M. Methodologies of the distributed systems buildup. Computational Technologies: Special Issue. 2006. Vol. 11. P. 3-17. (in Russian)
4. O.L. Zhyzhymov, A.M. Fedotov, O.A. Fedotova. Building a generic model of information system for working with documents on the scientific heritage // Vestnik. Novosibirsk State University. Series: Information Technologies. – 2012. – Vol.10. – № 3. – P.5-14. (in Russian)
5. DSpace: an open source solution for accessing, managing and preserving scholarly works // [online] / MIT Libraries; HP Labs. – 2007. <http://www.dspace.org/> (in English)
6. Reference Model for an Open Archival Information System (OAIS), Recommended Practice, Issue 2, CCSDS 650.0-M-2 (Magenta Book) June 2012. (in English) Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
7. 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 (in English) (<http://www.openarchives.org/>).

8. Reznichenko V.A., Proskudina G.Yu, Kudim K.A. Conceptual Model of Digital Library// Digital Libraries: Advanced Methods and Technologies, Digital Collections: Proceedings of the XI All-Russian Research Conference RCDL'2009. Petrozavodsk: KRC RAS, 2009. Pp. 23-31. (in Russian)
9. Lagoze C., & Van de Sompel, H. The open archives initiative: building a lowbarrier interoperability framework // In Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries. 2001. pp. 54-62. New York, NY, USA. (in English)
10. Sun S., Lamm L., & Boesch B. Handle System Overview (RFC 3650). Corporation for National Research Initiatives. 2003. Retrieved from <http://www.ietf.org/rfc/rfc3650.txt> (in English)
11. Directory of Open Access Repositories – OpenDOAR. (n.d.). Retrieved August 10, – 2013. (in English) From <http://opendoar.org/>
12. Goh, D. H.-L., Chua, A., Khoo, D. A., Khoo, E. B.-H., Mak, E. B.-T., & Ng, M. W.-M. // A checklist for evaluating open source digital library software. Online Information Review, – 2006. № 30 (4), pp. 360–379. (in English)
13. ANSI/NISO Z39.88-2004 (R2010) – The OpenURL Framework for Context-Sensitive Services. National Information Standards Organization. 2010 [2005]. ISBN 978-1-880124-61-1. (in English)
14. Kudim K.A., Proskudina G.Yu, Reznichenko V.A. Creation of scientific digital libraries using the system DSpace // Problems in programming – 2007. – №. 3. – С. 49-60. (in Russian)
15. Lewis S., de Castro P., & Jones, R. SWORD: Facilitating Deposit Scenarios. D-Lib Magazine, –2012. № 18 (1/2). (in English)
16. SWORD V2 Specifications. (n.d.). Retrieved August 10. 2013. (in English) from <http://swordapp.org/sword-v2/sword-v2-specifications/>
17. Pyrounakis G., Saidis K., Nikolaidou M., & Karakoidas, V. Introducing Pergamos// A Fedora-Based DL System Utilizing Digital Object Prototypes. In Research and Advanced Technology for Digital Libraries. 2006. pp. 500–503. Springer Berlin Heidelberg. (in English)
18. Library of Congress. (n.d.). METS: An Overview & Tutorial. (in English) Retrieved from <http://www.loc.gov/standards/mets/METSOverview.v2.html>
19. Tramboos S., Humma H., M Shafi S., & Gul, S. A study on the Open Source Digital Library Software: Special Reference to DSpace, EPrints and Greenstone // International Journal of Computer Applications. 2012. № 59 (16), pp. 1–9.
20. Reznichenko V.A., Proskudina G.Yu, Ovdii O.M. Creation digital libraries of periodicals with Greenstone // Digital library. 2005. V.8. – № 6. (in Russian) <http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2005/part6>
21. CERN Document Server. (n.d.). Retrieved August 10. 2013. from <http://cds.cern.ch/>
22. O.L. Zhyzhymov, A.M. Fedotov, Shokin Y.I. Platform ZooSPACE – providing access to heterogeneous distributed resources // Digital library. – 2014. – T.17. – № 2. – ISSN 1562-5419. (in Russian)
23. O. L. Zhizhimov, A. M. Fedotov, Yu. I. Shokin Technology platform for the mass integration of heterogeneous data // Vestnik. Novosibirsk State University. Series: Information Technologies – 2013. – T.11. – № 1. – С.24-41. – ISSN 1818-7900. (in Russian)