

Практические аспекты экспорта библиографических записей ИРБИС с целью их использования внешними интеграторами

К. А. Колосов

*Государственная научно-техническая библиотека России,
Москва, Россия*

В процессе деятельности библиотек все чаще возникает задача выгрузки больших массивов библиографических описаний из системы автоматизации библиотек ИРБИС в один или несколько файлов в коммуникативном формате с целью дальнейшего использования этих данных внешними интеграторами. Так, ГПНТБ России передала метаданные, выгруженные из баз данных ИРБИС, следующим внешним интеграторам:

- Ebsco для включения в информационный массив портала EBSCO Discovery Solutions;
- ExLibris для включения в информационный массив портала Primo Central Index;
- Google для поиска на портале Google Scholar;
- ProQuest для включения в информационный массив портала Summon.

Внешние интеграторы производят индексацию полученных метаданных и включают их в свои массивы документов, которые становятся доступными для поиска как пользователям на web-порталах этих интеграторов, так и на портале библиотеки, предоставившей метаданные. Передаваемые метаданные могут быть следующих видов:

- библиографические описания, выгруженные из электронного каталога библиотеки;
- библиографические описания полнотекстовых электронных документов;
- список электронных периодических изданий, на которые подписана библиотека.

Метаданные могут передаваться интеграторам в формате MARC21 и представлении ISO 2709, в формате MARCXML, в формате Dublin Core. Передача метаданных может осуществляться выгрузкой сформированных файлов на FTP сервер интегратора (ProQuest, Ebsco), размещением файлов на WEB сервере библиотеки в разделе, доступным для робота интегратора (Google), размещением метаданных в электронном репозитории с последующим их заимствованием интегратором по протоколу OAI-PMH (ExLibris).

Рассмотрим требования интеграторов к метаданным в формате MARC21, не поддерживаемые в базовой конфигурации ИРБИС, и вариант решения ГПНТБ России для их конвертирования в требуемый формат. Для связи передаваемых метаданных с исходными записями в электронном каталоге библиотеки в Интернет, который для библиотек, использующих ИРБИС, базируется на платформе WEB ИРБИС или J-ИРБИС, интегратор требует:

- Наличия правильно сформированного поля 001 формата MARC21 – уникальный идентификатор записи.
- Наличия поля, содержащего обратную ссылку в формате `http://<путь к записи в каталоге ОПАС библиотеки>`, которая должна отослать пользователя к исходной записи в электронном каталоге библиотеки.

Для формирования обратной ссылки на запись в каталоге ОПАС библиотеки в решении ГПНТБ России используется поле 903 ИРБИС – Шифр документа в БД, которое при правильном его заполнении является уникальным. Это же поле используется и при формировании поля 001 формата MARC21. Для обратной ссылки на запись в каталоге ОПАС в формате MARC21 не предусмотрено поля из обязательного блока полей, поэтому для этой цели требуется использовать какое-либо поле из блока локального использования (поля 9xx). В разработанном решении для этой цели используется поле 957 и подполе ^u. Содержимое этого поля формируется из следующих составляющих:

базовый адрес (URL) WEB ИРБИС библиотеки & префикс поиска по шифру документа & содержимое поля 903 ИРБИС & постфикс.

Для WEB ИРБИС это поле формируется следующим образом:

`http://<адрес web-сервера библиотеки>/cgi-bin/cgiirbis_64.exe?S21CNR=& S21STN=1&S21CNR=20&21COM=S&I21DBN=IBIS&P21DBN=IBIS&`

S21ALL=<.>I=значение подполя 903<.>&S21FMT=fullwebr&Z21ID=

Например, для записи каталога ГПНТБ России, поле 957 в формате MARC21 может иметь следующий вид:

```
#957: ^u http://library.gpntb.ru/cgi/irbis64r_simple/site/cgiirbis_64.exe?S21CNR=& S21STN=1&S21CNR=20&C21COM=S&I21DBN=IBIS&P21DBN=IBIS&S21ALL=%3C.%3EИ=06929000021%3C.%3E&S21FMT=fullwebr&Z21ID=
```

здесь %3C и %3E соответствуют UTF представлению символов “< “и “>”. Для этого же примера записи поле 001 формата MARC21 будет иметь вид:

```
#001: ГПНТБ России|IBIS|06924000021
```

Формировать поле 957 можно разными способами, например с помощью глобальной коррективы. В ГПНТБ России для этой цели используется способ конвертирования выгруженных в текстовый формат записей: база данных выгружается в текстовом представлении внутреннего формата ИРБИС, запускается конвертор, использующий заданный префикс WEB ИБРИС, для добавления к каждой библиографической записи поля 957, которое формируется по изложенным выше правилам. Полученный в результате конвертирования файл загружается в новую базу данных ИРБИС, после чего производится выгрузка всех записей с преобразованием в формат MARC21 (ISO 2709). Формирование словаря для целей выгрузки записей не требуется и весь процесс конвертирования и выгрузки занимает немного времени. В файле SMARCEW.fst, используемом для выгрузки записей с преобразованием в формат MARC21, сделаны следующие изменения:

```
для поля 001: 1 0 &unifor('Korg.mnu|2'),'&unifor('+D'),'v903
```

```
для поля 957: 957 0 " " ^u|v957 ^u
```

Конвертор, добавляющий поля 957 к записям ИРБИС в текстовом представлении, выложен на форум пользователей ИРБИС.

Приведенная технология может быть использована для выгрузки записей электронного каталога библиотеки с целью их представления на порталах EBSCO Discovery Solutions и Summon. Для передачи метаданных с их целью интеграции с массивом портала Primo Central Index используется иная технология, основанная на промежуточном использовании электронного репозитория. Перенос библиографических записей из базы данных ИРБИС в электронный репозиторий можно осуществить с использованием программного пакета ZMARCO, который обращается к серверу Z64 ИРБИС по протоколу Z39.50. Для правильной работы ZMARCO необходимо соблюдения следующих правил в записях базы данных ИРБИС, представленных в коммуникативном формате MARC21:

- Поле 001 должно содержать уникальный номер записи, по которому сервер Z64 будет осуществлять ее поиск и извлечение с использованием поискового атрибута 12 (Local number) набора bib-1.
- Поле 005 должно содержать дату и время последнего изменения записи и быть в корректном представлении: *уууymmdd*.
- Поле 008 в позициях 00-05 должно содержать дату ввода записи *уymmdd*, а в позициях 07-10 должно содержать дату публикации.
- Поле 856 в подполе ^u должно содержать обратную ссылку на запись в каталоге OPAC, кроме случая работы с каталогом полнотекстовых документов, когда в этом подполе содержится URL полного текста.

Процесс размещения метаданных в промежуточном электронном репозитории включает следующие технологические этапы:

- Данные БД ИРБИС выгружаются в файл в текстовом формате и обрабатываются конвертором.
- Создается новая БД ИРБИС, в которую загружаются записи после конвертирования, актуализируется словарь.
- Новая БД ИРБИС подключается к серверу Z64.
- Запускается программа ZMARCOoperator, формирующая список записей в БД MS Access. В качестве параметров указываются даты публикации документов: от года N1– до года N2.

- Создается виртуальный каталог сервера IIS, в качестве физического пути указывается директория ZMARCO.

- Редактируется файл global.asa в директории ZMARCO, в котором указывается адрес сервера Z64 и имя базы данных выгружаемой базы данных ИРБИС.

- После этого провайдер OAI-PMH готов обрабатывать запросы клиента OAI/PMH.

Пример ответа провайдера OAI-PMH можно посмотреть по ссылке: http://193.233.14.6:8000/irbis/oai.asp?verb=ListRecords&metadataPrefix=oai_dc.

Однако для практического использования данный вариант провайдера OAI-PMH не удобен, так как он работает медленно и подвержен сбоям при передаче данных, связанных со сложной цепочкой обработки запросов сервером Z64. Поэтому для целей передачи метаданных в Primo Central Index использовался дополнительный репозиторий на базе программного обеспечения Open Harvester Systems (<http://pkp.sfu.ca>). В данный репозиторий были последовательно загружены метаданные, получаемые через провайдер OAI-PMH ZMARCO. Поскольку репозиторий Open Harvester Systems для хранения метаданных использует СУБД MySQL, время, требуемое для извлечения метаданных из него существенно меньше, чем при работе с ZMARCO. Пример пользовательского интерфейса репозитория Open Harvester Systems в ГПНТБ России, используемого для передачи метаданных в Primo Central Index доступен по адресу: <http://ellib.gpntb.ru/ohs>. Пример обработки запросов OAI-PMH этим репозиторием доступен по адресу: <http://ellib.gpntb.ru/ohs/index.php/oai>.

Еще одной задачей, возникающей при работе библиотеки с внешними интеграторами, является формирование списка подписных электронных изданий в формате XML согласно требованиям разработчиков портала Google Scholar, изложенным в документе: http://scholar.google.ru/intl/ru/scholar/institutional_holdings.xml. Этот список может быть использован для отображения сведений о подписных электронных изданиях библиотеки при поиске статей на портале Академии Google (<http://scholar.google.ru>), а также на порталах интеграторов EBSCO Discovery Solutions, Primo, Summon и других. Для формирования данного списка использовался программный пакет, разработанный по проекту ГПНТБ России с НЭИКОН по выработке рекомендаций российским организациям для практического использования в системе управления внешним контентом. На рис. 1 представлен интерфейс подсистемы управления списком электронных подписных изданий библиотеки, а на рис. 2 приведен пример редактирования сведений о выбранном электронном издании. Данный программный пакет разработан на языке PHP и использует базу данных на основе СУБД MySQL. В качестве источника сведений о подписных периодических изданиях использовалась база данных подписки ГПНТБ России, содержащая наибольший среди российских библиотек список электронных периодических изданий. Библиотека, которая будет использовать разработанный программный пакет, отмечает в исходном списке лишь те издания, на которые у нее оформлена подписка. С учетом того, что в поисковом интерфейсе предусмотрен поиск по издательству, процесс отбора требуемых изданий из базового списка не занимает много времени. Выгрузка данных о подписных электронных изданиях библиотеки в формате, совместимом с Google Scholar, осуществляется нажатием соответствующего пункта меню программного пакета. На рисунке 3 приведен фрагмент выгруженных данных в формате XML для некоторых изданий, представленных в списке на рисунке 1.

Список подписных электронных изданий:

Добавить описание

Заглавие:

ISSN:

Издательство:

Название журнала, ISSN	Наличие подписки
<u>A P L Organic Electronics and Photonics</u> issn=1941-420X	
<u>A Radio</u> issn=1211-3557	
<u>A3 Bau</u> issn=-	
<u>A3 Volt</u> issn=U006-6656	
<u>AA Files</u> issn=0261-6823	
<u>Aandrijftechniek</u> issn=0165-5108	
<u>AAPA Asphalt Review</u> issn=0727-0003	
<u>AAPG Bulletin</u> issn=0149-1423	
<u>AATCC Review</u> issn=1532-8813	
<u>AATCC Review</u> issn=1532-8813	
<u>Abacus</u> issn=0001-3072	
<u>ABB Review</u> issn=1013-3119	
<u>ABI Technik</u> issn=0720-6763	
<u>Abstract and Applied Analysis</u> issn=1085-3375	
<u>Abstract Bulletin of Paper Science and Technology</u> issn=1523-388X	
<u>Abstracts in New Technologies and Engineering</u> issn=1367-9899	
<u>Abstracts in New Technologies and Engineering</u> issn=1367-9899	
<u>Abstracts Papers Presented American Mathematical Society</u> issn=0192-5857	
<u>Abstracts of the American Mathematical Society</u> issn=0002-9904	

Рис 1. Интерфейс подсистемы управления списком электронных подписных изданий библиотеки

Редактирование записи:

Название журнала:	A Radio
Место издания:	Praha
Издательство:	AMARO spol.s.r.o
Издается, начиная с года:	1998
Издавался до года:	
Электронная версия доступна с года:	
Электронная версия доступна до года:	
ISSN:	12113557
Страна:	CZ
Язык:	
Шифр журнала в базе данных:	W12600
Примечание:	
Наличие подписки: <input type="checkbox"/>	

Рис. 2. Интерфейс редактирования сведений об электронном издании

```
<?xml version="1.0" encoding="UTF-8" standalone="yes" ?>
<institutional_holdings><item type="electronic">
<title>A3 Bau</title>
<coverage><from><year>1982</year></from>
<to/></coverage></item>
<item type="electronic">
<title>A3 Volt</title>
<issn>U006-6656</issn>
</item>
<item type="electronic">
<title>AA Files</title>
<issn>0261-6823</issn>
<coverage><from><year>1981</year></from>
<to/></coverage></item>
<item type="electronic">
<title>Aandrijftechniek</title>
<issn>0165-5108</issn>
<coverage><from><year>1978</year></from>
<to/></coverage></item>
<item type="electronic">
<title>AAPA Asphalt Review</title>
<issn>0727-0003</issn>
<coverage><from><year>1982</year></from>
<to/></coverage></item>
<item type="electronic">
<title>AAPG Bulletin</title>
```

```

<issn>0149-1423</issn>
<coverage><from><year>1991</year></from>
<to/></coverage></item>
<item type="electronic">
<title>AATCC Review</title>
<issn>1532-8813</issn>
<coverage><from><year>2001</year></from>
<to/></coverage></item>
<item type="electronic">
<title>Abacus</title>
<issn>0001-3072</issn>
<coverage><from><year>1965</year></from>
<to/></coverage></item>
<item type="electronic">
<title>ABB Review</title>
<issn>1013-3119</issn>
</item>
<item type="electronic">
<title>ABI Technik</title>
<issn>0720-6763</issn>
<coverage><from><year>1980</year></from>
<to/></coverage></item>
<item type="electronic">
<title>Abstract and Applied Analysis</title>
<issn>1085-3375</issn>
<coverage><from><year>1996</year></from>
<to/></coverage></item>
<item type="electronic">
<title>Abstract Bulletin of Paper Science and Technology</title>
<issn>1523-388X</issn>
<coverage><from><year>1930</year></from>
<to/></coverage></item>

```

Рис. 3. Фрагмент данных об электронной подписке библиотеки в формате, совместимом с Google Scholar

Список дополнительных источников:

1. Колосов К. А. Интеграция ресурсов электронных библиотек, доступных по протоколу Z39.50 и электронных архивов, использующих протокол OAI-PMH [электронный ресурс] / К. А. Колосов. – Режим доступа: http://www.aselibrary.ru/conference/conference43/conference432199/_video/2220/. – Загл. с экрана.
2. Колосов К. А. Система управления каталогом электронных подписных изданий библиотеки с использованием архива НЭИКОН. Российская ассоциация электронных библиотек. [электронный ресурс] / К.А. Колосов . – Режим доступа: <http://aselibrary.ru/datadocs/pr/201306/Kolosov.pps>. – Загл. с экрана.