

Формат RUSMARC в библиотечных системах

Национальная служба развития системы форматов RUSMARC на протяжении нескольких лет проводит тестирование библиотечных систем на предмет обеспечения поддержки формата. В настоящее время свидетельства о соответствии имеют программные продукты: OPAC-Global (ДИТ-М), АБИС «Руслан» (ООО «Открытые библиотечные системы»), модуль «Каталогизатор» автоматизированной информационно-библиотечной системы «ИнфоКомм» (ООО «ИнфоКомм»), программный продукт «Книжный заказ» (Некоммерческий Фонд поддержки книгоиздания, образования и новых информационных технологий «Пушкинская библиотека»), АБИС «А-Elite» (ЗАО «ГИПЕР-GIPER»), «Модуль «Каталогизатор» системы поточной обработки информации «Элар-РК» (ООО «ПроСофт-М»), АБИС «АС Библиотека 3» (ООО Информационно-аналитические системы), АБИС «Академия+» (Центр автоматизированных технологий «Ростехноком»), АБИС «МАРК-SQL 1.9» (НПО «Информсистема»). Требования, которые предъявлялись системам при проведении тестирования с точки зрения поддержки формата, все это время оставались внутренними; сейчас мы посчитали необходимым поговорить о том, что же означает «поддержка формата», и проанализировать, как существующие системы решают задачи, связанные с форматом, прежде всего – библиографическим.

В описании любой библиотечной системы, как правило, обязательно заявляется: система полностью совместима с форматами RUSMARC, MARC21, UNIMARC и т.д. При этом какого-либо стандарта в этой области не существует, как нет и общепринятого понимания поддержки формата. Смысл, который вкладывается разработчиками систем в это положение, может быть совершенно различным.

В 2002 г. Американской национальной организацией стандартов опубликован документ «*Request for Proposal Writer's Guide to Standards for Library Systems*» – Руководство для составителей запроса предложений применительно к стандартам для библиотечных систем. Руководство предназначено прежде всего не для разработчиков систем, а для заказчиков – библиотек: в нем по сути фиксируются условия задачи со стороны заказчика, которые придется решать разработчику библиотечной системы, и, в частности, требования, которым должна удовлетворять система с точки зрения поддержки библиотечных и коммуникационных стандартов. В разделе «Библиографические форматы» отмечается, что библиографические данные представляют собой центральный компонент АБИС, именно они формируют основу ЭК и процессов кооперированной каталогизации. Так или иначе библиографические данные используются всеми функциональными модулями интегрированных библиотечных систем.

В соответствии с Руководством, проектное задание на разработку АБИС должно включать следующие общие положения:

- система должна кодировать БЗ в MARC-формате без ограничения длины записи (кроме ограничений, определяемых форматом);
- система должна обеспечивать импорт и экспорт БЗ в формате без дополнительного вмешательства разработчика, с полным сохранением определителей содержания (меток полей, индикаторов, идентификаторов подполей) и сохранением порядка полей в записи;
- система должна обеспечивать вывод БЗ с определителями содержания на рабочие станции пользователей.

Система должна предусматривать поддержку всего набора определителей содержания, даже если та или иная библиотека предполагает использовать только часть из них; должна соблюдаться структура записи, определенная в формате; должны быть реализованы механизмы валидации определителей содержания и некоторых контролируемых значений (например, кодов языков и кодов стран).

Далее в Руководстве отмечается, что структурированное представление элементов данных БЗ обеспечивает основу для создания мощной поисковой системы. При оценке библиотечной системы необходимо, в том числе, определить, насколько эффективно используются эти возможности.

Еще одно обязательное требование, зафиксированное в Руководстве, – динамичность системы, ее возможность реагировать на развитие формата и изменения, которые регулярно вносятся в формат.

Как правило, библиотечная система обеспечивает целый ряд функций, связанных с MARC-записями: прежде всего, это импорт и экспорт записей, каталогизация, поиск информации и вывод ее для пользователя. Развитие общих положений Руководства применительно к этим процессам и формату RUSMARC приводит к следующим требованиям.

I. Общие требования

1. Отсутствие ограничений на длину записи, полей, подполей (за исключением ограничений, регламентированных в формате)
2. Поддержка всех существующих и ранее использовавшихся меток полей, идентификаторов подполей, индикаторов, разделителей, определенных в формате
3. Поддержка повторяющихся полей и подполей в соответствии с форматом
4. Поддержка техники встроенных полей
5. Реализация связей между записями
6. Сохранение порядка полей, определенного каталогизатором при вводе

7. Сохранение порядка подполей, определенных каталогизатором при вводе, для тех полей, заполнение которых должно осуществляться в соответствии с правилами каталогизации или индексирования
8. Проверка правильности записи при вводе и редактировании, с генерацией соответствующих сообщений об ошибках
9. Наличие справочников контролируемых данных (прежде всего, кодированных) и контроль их использования в записях
10. Возможность работы с неполными записями (например, на этапе заказа)
11. Возможность внесения изменений в системе в случае изменений формата

II. Поиск

Для обеспечения поиска в массиве MARC-записей система должна обеспечивать:

1. Возможность индексирования до уровня подполей и индикаторов (с выбором полей и подполей для индексирования)
2. Возможность индексирования ключевыми словами (для выбранных полей, подполей, значений индикаторов)
3. Возможность настройки индексов – т.е. указания, данные каких полей и каким образом используются для формирования словарей

III. Импорт и экспорт записей

Возможность импорта и экспорта корректных MARC-записей является обязательным условием, без которого невозможно говорить об эффективной работе в корпоративной среде.

1. Импорт и экспорт записей в MARC-формате должен осуществляться без дополнительного вмешательства разработчика с сохранением всех определителей содержания.
2. Экспорт записей должен осуществляться в соответствии со стандартом ISO 2709 ISO 25577 (MarcXchange)
3. Должна быть предусмотрена возможность использования локальных полей (полей блока 9--), которые определяются соглашениями между участниками обмена
4. Необходим контроль дублетов при импорте. Параметры выявления дублетов и действия при их обнаружении должны определяться библиотекой-пользователем АБИС (поля, используемые для сравнения, возможность слияния дублетов или отказа от него, определение полей, защищенных от изменения при слиянии)
5. Должна допускаться возможность импорта неполных или кратких записей
6. Если RUSMARC является только коммуникативным форматом – необходимо корректное преобразование записи из внутреннего формата в RUSMARC (при импорте) и обратно (при экспорте)
7. Возможность работы с авторитетными файлами, в том числе – национальными

8. Контроль правильности записи

Контроль правильности записи при вводе и загрузке проводится для проверки структуры и содержания записи с точки зрения синтаксиса и семантических правил. Условия проверки можно разделить на две группы: (1) параметры проверки структуры и содержания отдельных элементов записи (поля, подполя, индикаторы), и (2) параметры перекрестной проверки структуры и содержания взаимозависимых элементов записи. Первая из этих групп включает следующие уровни проверки:

1. Соответствие меток полей и идентификаторов подполей спецификациям формата
2. Наличие обязательных полей в записи и подполей в полях (для формата RUSMARC – наличие полей 001, 100, 200, 801)
3. Соответствие повторяемости элементов данных (полей и подполей) условиям формата
4. Корректность использования значений индикаторов
5. Длина полей и подполей фиксированной длины
6. Корректность данных в полях фиксированной длины – соответствие типу данных, определенному в формате (например, в поле 100 позиции 8-11 должны содержать 4 цифры, обозначающие год, либо пробел)
7. Соответствие значений в полях кодированной информации спискам кодов, включенным в формат

Вторая группа – перекрестные проверки, – это проверки наличия и значений элементов данных, которые зависят от наличия / содержания других элементов записи. В частности, к этой группе относится проверка наличия в записи полей, имеющих статус «обязательное» для определенных типов данных (например, поле 230 является обязательным для электронных ресурсов, поля 120, 123, 206 – для картографических документов). Еще несколько примеров:

- соответствие значений позиции 8 в поле 100 (тип даты публикации) и позиций 9-16 (дата публикации 1 и дата публикации 2)
- обязательное наличие подполя 200\$z (язык параллельного заглавия) при наличии подполя 200\$d (параллельное заглавие)
- наличие в записи не более одного из полей 700, 710 или 720, определяющих первичную ответственность за содержание документа, и т.д.

Очень важно, чтобы контроль осуществлялся не только применительно к основным полям записи, но и ко встроенным полям в блоке связанных записей (блоке 4--). Дополнительное условие проверки в блоке 4-- – наличие в базе данных записи, номер которой указан во встроенном поле 001 – в соответствии с форматом, «если поле связи содержит контрольный номер связываемой записи, то эта запись включается в тот же файл».

Такой контроль следует применять (хотя и с разной степенью строгости) при сохранении записи в процессе создания и редактирования и при импорте записей из внешних источников. Выявленные ошибки могут быть исправлены сразу же (при каталогизации), либо, если они

выявлены при загрузке – включены в протокол проверки для внесения необходимых корректировок в записи позднее. Должна быть предусмотрена возможность настройки условий контроля записей при внесении изменений в формат.

Что зачастую происходит на практике в работающих библиотечных системах?

- Набор полей и подполей в системах часто оказывается неполным – он включает меньше полей / подполей, чем предусмотрено форматом. Ограниченный набор данных может определяться целевой категорией пользователей системы. Следует, однако, помнить, что если для сиюминутных потребностей ограниченный набор определителей содержания достаточен, в перспективе это может создать трудности при обмене записями более высокого уровня полноты – часть информации при загрузке может быть утрачена.
- Контроль записи при загрузке часто или отсутствует вообще, или осуществляется в недостаточной мере.
- В режиме создания и редактирования записи контроль, как правило, предусмотрен, но также не всегда достаточный – не всегда контролируется повторяемость полей и подполей, наличие обязательных полей и подполей, соответствие кодированных данных, реже всего проверяются взаимозависимости между элементами записей.
- Структура встроенных полей, как правило, вообще не подвергается контролю.
- Основная задача по созданию качественных корректных записей часто ложится на каталогизаторов и во многом определяется именно их квалификацией, вниманием, добросовестностью. Иногда – даже в тех случаях, когда могут использоваться значения по умолчанию (например, пробелы в позициях индикаторов, которые не определены), либо в случае, когда значения могут формироваться автоматически (например, коды наборов символов в поле 100 при выгрузке записей).
- Не всегда сохраняется порядок полей и подполей в записи. В результате может происходить искажение данных в тех полях, где порядок подполей определяется правилами каталогизации и индексирования – прежде всего, полях 200, 210, 225, полях предметных рубрик в блоке 6--.
- Это, в свою очередь, может привести к невозможности работы в корпоративной среде с национальными авторитетными файлами.
- Большая группа ошибок связана с некорректным преобразованием записей при загрузке / выгрузке; такие ошибки характерны для систем, в которых RUSMARC является только коммуникативным форматом (но не используется как внутренний формат). Больше всего проблем возникает с блоком связи записей 4--, в частности, с поддержкой многоуровневых записей.
- Еще одна значительная группа ошибок является результатом стремления разработчиков облегчить работу каталогизаторам. Для этого применяется принцип однократного ввода, с тем, чтобы в последующем необходимые элементы MARC-записи формировались на основе

введенных один раз данных. Самый частый пример – формирование подполей 200\$f и \$g на основе полей блока 7--. Однако, данное поле 200 соответствует области заглавия и сведений об ответственности ГОСТ 7.1-2003, и данные в нем должны вводиться в соответствии с правилами каталогизации в том виде, как они представлены в источнике описания (на титульном листе и т.д.). В отличие от поля 200, поля 7-- блока содержат нормированные точки доступа, форма которых может отличаться от формы, приведенной на титульном листе.

Например, в исходной записи:

```
200 1#$aОбщевоинские уставы Вооруженных сил Российской Федерации
...
710 01$3RU\NLR\auth\884673$aРоссийская Федерация$bВоенные уставы и
наставления
```

После загрузки:

```
200 1#$aОбщевоинские уставы Вооруженных сил Российской Федерации
    $fРоссийская Федерация, Военные уставы и наставления
...
710 02$aРоссийская Федерация$bВоенные уставы и наставления
```

Возникает противоречие: с одной стороны, разработчики систем в своей заботе о каталогизаторах часть работы берут на себя (при этом не всегда делая ее правильно), с другой стороны – перекалывают на каталогизаторов техническую работу, которую логичнее было бы поручить компьютеру.

- Следующая группа ошибок – некорректная обработка значений индикаторов.

Например, в поле 464 (аналитический уровень) индикатор указывает необходимость формирования примечания о содержании. В одной из систем, прошедших тестирование в Национальной службе RUSMARC, поле 327 (примечание о содержании) формируется вне зависимости от этого значения. В результате запись может содержать два поля 327 с аналогичными данными.

Например, в исходной записи:

```
327 0#$aВ кн. также: У. Черчилль в зеркале цитат и высказываний
464 #0$12001#$aУинстон Черчилль в зеркале цитат и высказываний
```

После загрузки:

```
327 1#$aСодержание : Уинстон Черчилль в зеркале цитат и высказываний
327 0#$aВ кн. также: У. Черчилль в зеркале цитат и высказываний
464 #0$12001#$aУинстон Черчилль в зеркале цитат и высказываний
```

- При выгрузке записей наборы символов, указываемые в поле 100 (поз. симв. 26-29 и 30-33), не соответствуют реальным наборам символов, используемым в выгружаемых записях. Типичная ситуация: например, загружается запись в кодировке Windows 1251 (в поле 100 / 26-29 указаны коды 0189). При выгрузке этой записи в другой кодировке (например, UTF-8) поле 100 остается без изменений, хотя в соответствии с форматом в этом случае позиции 100/26-29 должны содержать коды 50##.
- Еще одна характерная особенность некоторых АБИС является оборотной стороной стремления разработчиков обеспечить гибкость системы, возможность ее адаптации под потребности каждого конкретного пользователя. Часто оказывается стерта грань между возможностями настройки системы, которые предоставляются пользователю, и исправлением пользователем неточностей (а иногда – и ошибок), допущенных разработчиками.

С самого своего появления формат всегда выполнял роль посредника между правилами каталогизации и разработчиками программного обеспечения для библиотек – помимо чисто каталогизационных данных, формат включает и элементы, не предусмотренные правилами каталогизации. Это могут быть кодированные данные, предназначенные для автоматической обработки (например, коды языков). Кроме того, некоторые элементы данных БЗ в формате представлены в кодированном виде для обеспечения автоматической обработки (например, дата публикации в поле 100). Развитие форматов всегда направлено на то, чтобы удовлетворять меняющиеся потребности как каталогизационного сообщества, так и разработчиков библиотечных систем. И сейчас можно говорить о том, что каталогизатор не является единственным создателем библиотечного каталога; в равной степени каталог – это продукт усилий программистов, направленных на максимальное использование потенциала записей.

Но – насколько эффективно используются в системах данные, содержащиеся в MARC-записях?

К сожалению, можно отметить некоторое пренебрежение со стороны разработчиков АБИС к элементам данных, которые, хоть и не являются обязательными, но при правильном и полном использовании оказываются вовсе не бессмысленными. В частности, это относится к кодированной информации. Часто оказывается, что даже обязательные элементы данных в записи присутствуют, но лишь формально – заполняются либо значениями по умолчанию, либо символами – заполнителями. Да, это не противоречит формату; да, это облегчает работу с системой для пользователей. Но полноценно работать такая запись не будет. Кроме того, в корпоративной среде эффективная работа возможна при сравнимом уровне качества записей; разный уровень полноты записей, поставляемых участниками, может привести к потерям при поиске. В качестве примера можно привести коды читательского назначения (поле 100, поз. симв. 17-19). Эти позиции являются факультативными, но отсутствие кодов в записи означает, что при поиске, например, литературы для дошкольников (который мог бы проводиться на основе кода в

поле 100), такие записи будут просто проигнорированы. Особенно это актуально в сводных каталогах, участники которого работают в разных системах.

Еще один пример – код таблиц классификации в поле 686 (\$v). Известно, что разные библиотеки работают по разным вариантам таблиц Библиотечно-библиографической классификации – для научных библиотек, массовых, детских и т. д. Подполе 686\$v имеет статус факультативного, но его отсутствие или не учитывание может привести к ошибкам при поиске и при выводе записи: так, при выводе записи знаком «+» ошибочно соединяются индексы разных таблиц ББК: 47.2 + п722.42(92) + п722.42(94).

Еще пример: данные, необходимые для обработки записи. В условиях работы в корпоративной среде нелишне иметь информацию об источнике записи и дате ее редактирования. Для этого в формате используются поля 005, 035, 801. Поле 005 содержит дату последнего редактирования, 035 – идентификатор записи в системе, из которой она была получена, 801 – даты создания и редактирования записи. Если все поля используются полностью, мы можем получить полную историю работы с записью. Однако часто факультативные поля 005 и 035 не используются вообще, а 801 поле используется только для отражения даты создания записи. Здесь уже нельзя прикрыться заботой о пользователе. Эти поля могут формироваться не каталогизатором, а системой, и их неиспользование связано только с нежеланием производителей программного обеспечения.

Далеко не всегда MARC-данные в полной мере используются для организации эффективного поиска. При правильном использовании MARC-записи могут обеспечить и организацию фасетной навигации, и ранжирование записей при выводе в зависимости от релевантности с приписыванием разных весов поисковым терминам в зависимости от того, в каком поле записи они встречаются, и т. д. Но это – тема отдельного разговора, выходящая за рамки данного сообщения.

Все это приводит нас к заключению: несмотря на то, что, как правило, производители каждой библиотечной системы заявляют о полной поддержке RUSMARC, это заявление далеко не всегда соответствует действительности; кроме того, существующие системы не всегда используют потенциал формата в должной степени. В то же время можно по-разному относиться к формату, рассуждать о его недостатках, но, тем не менее, он по-прежнему остается общим языком, обеспечивающим коммуникацию между системами, и реальность такова, что таковым формат останется еще долго. А насколько эффективно этот язык используется – зависит от нас.