

## **MARC-форматы в современной информационной среде**

Говоря о месте MARC-форматов в современной информационной среде, необходимо учитывать ряд общих факторов, оказывающих влияние на формирование и функционирование информации. Среди основных факторов – развитие систем метаданных для описания ресурсов, рост числа создателей информационного контента, резкое увеличение доли электронных ресурсов, включая полнотекстовые базы данных, развитие Интернет-технологий, обеспечивающих более полный, удобный и оперативный поиск и тем самым составляющих серьезную конкуренцию библиотечным каталогам.

В то же время развитие библиотек во многом остается инерционным, а библиотечный мир – во многом замкнутым. Данные, которые хранятся в каталогах и базах данных отдельных библиотек, часто оказываются закрытыми для веб-приложений, и не учитываются при поиске, выполняемом поисковыми машинами, даже если сам каталог может быть доступен для удаленных пользователей. При этом мы по-прежнему не готовы признать, что библиотеки все больше теряют центральную роль основного хранителя и поставщика информации.

Библиографическая инфраструктура, существующая сейчас в библиотечном мире, часто не рассчитана на поддержку нескольких схем метаданных. Основной, а во многих случаях и единственной схемой метаданных, используемой в библиотеке, остается MARC-формат. Какие же еще элементы входят в эту инфраструктуру, и как они соотносятся с MARC-форматом? Прежде всего, это различные схемы метаданных, такие как Dublin Core, MODS, METS.

### MARC и другие стандарты метаданных

Схема **Dublin Core (DC)** создавалась как простой набор элементов данных для описания документов и других объектов в Интернете. Dublin Core состоит из 15 основных элементов, которые при необходимости могут дополняться квалификаторами. Благодаря своей компактности и простоте схема получила очень широкое распространение.

При разработке DC не предполагалось, что новая схема полностью заменит MARC, хотя такие мнения и высказывались. Действительно, схема DC более приспособлена к использованию поисковыми инструментами, однако записи DC не

обеспечивают такую полноту и детальность, как MARC-записи. В связи с этим возник вопрос – а действительно ли необходимо подробное описание (MARC или даже DC) для поиска интернет-ресурсов. Может, можно вообще ограничиться только полнотекстовым поиском, который обеспечивают поисковые машины в Интернете? Дискуссии привели к предложению определить 4 уровня каталогизации электронных ресурсов в зависимости от важности ресурса: полные MARC-записи, дополненные записи DC, простые записи DC, и использование только поиска по ключевым словам, предлагаемым поисковыми машинами. Уровень детальности описания должен определяться каталогизатором для каждого отдельного ресурса или категории ресурсов.

**MODS (Metadata Object Description Standard).** Схема MODS, разработанная Библиотекой Конгресса в 2002 г., представляет собой сокращенную, более «дружественную» для пользователя версию MARC – подмножество ключевых элементов данных MARC переведено в легко понимаемый XML-формат. Вместо трехзначных меток полей, абстрактных идентификаторов подполей используются понятные для пользователя, вербальные метки (например, «title» вместо «245»). Большая часть элементов данных фиксированной длины игнорируется. Определены новые элементы данных, например: «name», который включает и личное имя, и наименование организации, и может использоваться и в поле автора, и как часть предметной рубрики.

Хотя схема MODS создана на основе MARC21 и намного детальнее, чем DC, в ней намного меньше правил, чем в MARC. Как и в DC, нет обязательных полей, все поля могут повторяться.

Записи MODS часто используются в базах данных, которые включают смесь библиотечной каталогизации и библиографических данных, полученных из других источников.

**METS – (Metadata Encoding and Transmission Standard).** Этот формат метаданных используется цифровыми библиотеками и архивами. Задача METS – не «описание» в привычном для каталогизатора смысле этого слова. METS выполняет роль «обертки» и служит для того, чтобы собрать вместе все файлы, составляющие цифровой объект. В отличие от книги, которая представляет собой единый физический объект, цифровой объект может состоять из множества отдельных файлов (например, представляющих отдельные страницы). И, в отличие от книги, у цифрового объекта нет обложки или титульного листа, мы не можем просто перелистать страницы для того,

чтобы найти нужное место в книге. METS можно сравнить с переплетом, обложкой и навигационным инструментом для группы цифровых файлов. Кроме того, он включает техническую информацию, необходимую для понимания файлов и управления ими – формат файлов, технология сканирования и т.д. Описательные данные в METS не определены. Вместо этого допускается вставлять в записи METS любые дескриптивные метаданные, необходимые для описания материалов. Таким образом, записи METS запросто могут содержать дескриптивные данные в DC или MODS.

**ONIX (Online Information Exchange)** – XML-схема, разработанная в издательском сообществе в качестве международного стандарта для представления информации о книгах, периодических изданиях, видеопродукции в электронном виде.

Несмотря на то, что ONIX и MARC нацелены совершенно на разное использование данных, все-таки говорить об их полной несовместимости нельзя. Как минимум одна общая точка у них есть – это сам объект описания – с одним и тем же названием, одним и тем же автором, изданный одним издательством. Это дает возможность строить конверторы, таблицы соответствий между элементами данных разных форматов. В частности, такой конвертор существует в Библиотеке Конгресса.

#### MARC и XML

XML – это стандарт-контейнер, метаязык разметки, который обеспечивает простое структурное представление данных (*создание собственно языков разметки*), имеет весьма широкую область применения и играет все большую роль в информационном обмене.

С самого возникновения и начала распространения XML все чаще стало высказываться мнение о том, что MARC-формат окончательно отжил свое и должен уступить свое место XML. О преимуществах XML над MARC-форматами в последние годы не говорил только самый ленивый:

1. Гибкость: в нем нет фиксированного набора полей и определенного синтаксиса, это позволяет пользователям создать свои собственные языки. Эта гибкость обеспечивает легкую адаптируемость к стандартам каталогизации.
2. Отсутствие жестких ограничений на размер записи.
3. Возможность включения в запись других текстов, имиджей, звуковых файлов.
4. Широкое распространение вне библиотечного сообщества.

Однако со временем высказывания становились все менее и менее радикальными. Так, один из ярых противников MARCа, Рой Теннант в 2002 г. опубликовал статью с красноречивым названием «MARC должен умереть». Через 2

года он уже пишет: «Я полагал, что MARC (синтаксис записи MARC, элементы данных MARC, AACR2) слишком ограничены для того, чтобы удовлетворить современные потребности библиотек и их пользователей. Теперь я понимаю, что, имея сильную библиографическую инфраструктуру, мы можем использовать любой стандарт библиографических метаданных, в том числе MARC.» Еще через 2 года (2006): «...умереть должен не MARC и не AACR2 конкретно, несмотря на все их проблемы и недостатки, а наша абсолютная убежденность в том, что MARC и AACR2 – единственные возможные требования к библиографическим метаданным. Мы должны создать инфраструктуру, которая может работать с MARC наряду с другими стандартами метаданных. Мы должны ассимилировать MARC в более широкий, более богатый, более разнообразный набор инструментов, стандартов и протоколов.» Слухи о смерти MARC оказались сильно преувеличены.

Итак, говоря о взаимоотношениях MARC и XML, нельзя ставить вопрос «или/или». Библиотеки так много вложили и сил, и средств в MARC-записи, что было бы глупо просто отказаться от них. Гораздо логичнее максимально их использовать.

Разрабатываются различные версии XML-схем для MARC-форматов. В частности, версия библиотеки Конгресса, разработанная изначально для MARC21, послужила основой для разработки стандарта ISO 25577 – MarcXchange. Стандарт определяет требования к общей XML-схеме, для обмена как библиографическими записями, так и другими типами метаданных. Это «транспортная схема», которая может быть применена для любого MARC-формата. В стандарте не определяется содержание отдельных записей, не определяется значение конкретных меток, индикаторов, идентификаторов подполей. Описана общая структура, предназначенная для обеспечения обмена между системами обработки данных. Валидация содержания MARC-записей обеспечивается специальными программными средствами, ориентированными на конкретный MARC-формат.

Общая схема ISO 25577 не предусматривает использование встроенных полей, для UNIMARCa (соответственно RUSMARCa) схема требует соответствующих дополнений. «Тонкая» XML-схема (RUSMARXML/Slim) для библиографических записей в формате RUSMARC разработана Центром компьютерных технологий Уральского Государственного университета им. А. М. Горького. Схема, соответствующая ISO 25577 и дополненная для UNIMARC, разрабатывается Постоянным комитетом по UNIMARC при участии Национальной службы развития системы форматов RUSMARC.

## MARC и правила каталогизации

Хотя MARC-формат определяет и структуру, и семантику библиографической информации, все же прежде всего это стандарт структуры данных, а не стандарт содержания. Содержание записи регламентируются правилами каталогизации, системой предметизации и т.д.

В то же время есть ряд вопросов, трактовка которых в формате и в правилах каталогизации различается. Один из таких спорных вопросов – разделение между содержанием записи и носителем. У участников Сводного каталога библиотек России возникает много проблем при описанию электронных ресурсов, в частности, при идентификации описываемого ресурса как электронного / видеоматериала / звукозаписи.

В RUSMARC (как, собственно говоря, и в UNIMARC, и в MARC21) принят следующий подход: при описании ресурса, который сочетает характеристики разных видов материала, в записи отражаются характеристики каждого вида материала для описания всех аспектов ресурса, включая и содержание, и физический носитель, при этом основным является именно содержание, а потом уже – носитель. При определении типа записи, который указывается в маркере (по сути – идентификация ресурса как электронного, видео, аудио и т.д.) *отдается предпочтение коду вида каталогизируемого материала, а не коду вторичной физической формы (формат)*.

В Российских правилах каталогизации принимается во внимание как форма воплощения, так и содержание ресурса, но первичным оказывается носитель. Так, в гл.10 п.3. Общее обозначение материала: *«...предпочтение отдают обозначению физической формы, в которой представлен материал»*, и далее, в п.3.5: *«Если объект описания является особым видом документа (карты, ноты, изоматериалы и т.п.) или документом, расположенным на специфическом носителе (ЭР, микроформа и т.п.), то предпочтение отдают общему обозначению материала этого носителя»*.

Вообще, практика каталогизации нацелена прежде всего на документ как физический объект описания, а не на описание интеллектуального содержания. В результате во многих случаях создаются десятки записей на документы с идентичным содержанием. Эту ситуацию призвана разрешить модель FRBR (Functional Requirements to Bibliographic Data – Функциональные требования к библиографическим записям), которая ориентирована прежде всего на содержание, на поддержку связей между произведениями, создатели модели предполагают, что ее применение позволит пользователю проще ориентироваться в каталоге.

В настоящее время нельзя говорить, что модель FRBR инкорпорирована в MARC-формат или правила каталогизации, но работа в этом направлении ведется. Предпринимаются попытки реализовать модель на практике, для поиска в библиотечных системах.

Первая попытка применения модели FRBR в правилах каталогизации предпринята создателями RDA (Resource Description and Access – Описание ресурсов и доступ к ним). RDA – свод правил, который призван прийти на смену AACR и разрабатывается в качестве нового стандарта содержания для описания ресурсов и обеспечения доступа к ним.

Ключевые моменты создаваемого свода правил: (1) в основе RDA лежат концептуальные модели FRBR и FRAD (Functional Requirements to Authority Data – Функциональные требования к авторитетным данным); RDA ориентирован на объектно-ориентированную структуру баз данных; (2) проводится четкое различие между содержанием записи и ее представлением. Основным является именно содержание: «Задача – создать набор инструкций для записи данных, которые могут быть применены независимо от какой бы то ни было структуры или синтаксиса для хранения или вывода данных». Вывод (представление) записи будет рассмотрен в отдельном приложении, с указанием соответствий между элементами RDA и ISBD, для формирования привычного вывода библиографического описания (ISBD). Кроме того, отдавая должное широкому распространению MARC-форматов в библиотеках, предполагается включить приложение с инструкциями по представлению описательных элементов данных в MARC-формате.

В марте 2008 г. Комитетом по разработке RDA под эгидой Британской библиотеки, Библиотеки Конгресса, Библиотеки и архива Канады создана рабочая группа RDA/MARC, среди задач которой – разработать предложения для обсуждения с MARC-сообществом; определить, какие изменения необходимы в MARC21 для поддержки совместимости с RDA и эффективного обмена данными в будущем. В ближайших планах – обеспечить совместимость RDA и MARC-формата на уровне словарей.

\*\*\*

Таковы основные элементы информационной инфраструктуры. Для того, чтобы это была действительно единая инфраструктура, мы должны обеспечить реальное взаимодействие этих компонентов друг с другом. Кроме того, каждый компонент должен использоваться максимально эффективно.

Многие проблемы, которые, казалось бы, возникают при работе с MARC-записями, - это на самом деле проблемы не записей, а библиотечных систем. MARC-записи, накопленные за время использования формата, представляют собой огромный массив стандартизированных данных. Задача библиотечных систем – максимально использовать это стандартизированное, структурированное содержание.

Например, большинство элементов FRBR на самом деле включаются в записи уже на протяжении десятилетий, но только сейчас мы выяснили, что они были бы полезны для пользователя – помогли бы ему ориентироваться в каталоге.

Как еще элементы MARC-записи могут облегчить работу с каталогом, сделать ее более удобной и эффективной?

(а) Фасетная навигация на основе определителей содержания MARC – многомерный просмотр (*browsing*), отражающий связи между записями в массиве. Например, тематический поиск может дать тысячу записей. Получить более точный результат может уточнение по индексу, дате публикации, месту хранения и т.д. Если данные для формирования фасетов берутся из реальных записей – такое сужение никогда не приведет к нулевому результату.

(б) Ранжирование результатов поиска по релевантности (приписывание различных весов терминам запроса в зависимости от того, в каком поле найден термин)

(в) Различная сортировка результатов поиска по релевантности, дате издания, заглавию, автору, востребованности и популярности документа (насколько часто его ищут пользователи) и т.д.

Очевидно, что реализация этих функций возможна только при условии достаточной информации в записи.

В той или иной мере эти функции реализуются в таких проектах, как: FictionFinder (OCLC), RedLightGreen (проект Research Libraries Group, позднее вошел в каталог OCLC WorldCat, Endeca – web-интерфейс Библиотеки Университета Северной Каролины).

\*\*\*

Надо сказать, что значение информации, накопленной в MARC-записях, признается не только библиотечным сообществом. Со стороны поисковых служб Интернета предпринимаются усилия к интеграции традиционного библиотечного содержания и служб в своих поисковых машинах.

В качестве примеров можно назвать два проекта Google:

1. Google LibraryThing – веб-приложение для создания каталогов библиотек любого уровня, хранения и обмена библиографическими записями. Основная функция LibraryThing – импорт данных от книготорговых организаций и библиотек, при этом библиотеки поставляют записи в MARC-формате и DC.

2. Google BookSearch – проект, направленный на создание всеобъемлющего виртуального каталога для поиска всех книг на всех языках при участии издательств и библиотек. В мае 2008 г. подписано соглашение между Google и OCLC, в соответствии с которым, в частности, библиотеки-партнеры OCLC передают в BookSearch свои MARC-записи. Тем самым расширяются возможности доступа к библиотечным коллекциям через веб.

\*\*\*

Итак, современная библиографическая инфраструктура должна обеспечить быстрый и беспроблемный обмен различными типами данных. Требования, которым должна удовлетворять эта инфраструктура, многочисленны и разнообразны. Первичным является пользователь, его потребности и наша способность эти потребности удовлетворить:

- универсальность (многосторонность) – возможность работы с метаданными из различных источников, описывающими различные виды объектов. Мы должны иметь возможность принять запись от издателя (например, в ONIX), и использовать ее в качестве прототипа каталогизационной записи, или добавить информацию в запись, которая уже есть в нашем каталоге.
- расширяемость – возможность включения новых стандартов метаданных и использования любого стандарта по выбору
- открытость и прозрачность
- поддержка иерархических структур
- достаточный уровень детализации информации

MARC-формат остается одним из компонентов этой библиографической инфраструктуры, наряду с другими схемами метаданных. При этом каждый компонент библиографической инфраструктуры должен выполнять свои функции, обслуживать интересы и потребности «своей» категории пользователей.