

Заметки, которые родились в процессе подготовки к панельной дискуссии "MARC: заменить нельзя оставить", проходившей в рамках III Библиографического конгресса.

MARC: заменить нельзя оставить

За рубежом к обсуждению этого вопроса вплотную приступили лет 20 назад. Безусловно, отголоски споров, которые начались со знаменитого высказывания Роя Теннанта «MARC must die», доходили и до российских библиотек. В начале 2010-х гг. разработчики одной из систем, проходивших сертификацию в Национальной службе поддержки системы форматов RUSMARC (а Служба проводила сертификацию программных продуктов на предмет поддержки формата), получая сертификат, спросили – «А правда, что RUSMARCa скоро не будет?». Прошло без малого 20 лет, формат жив. Большею частью мы наблюдали за спорами со стороны; всплеск дискуссии о судьбе национального формата был в 2015-2016 году, в связи с закрытием Отдела форматов в Российской национальной библиотеке, на базе которого работала Национальная служба RUSMARC. В последнее время разговоры опять оживились – и к нашему сообществу приходит понимание необходимости поиска новых путей для сохранения места библиотек в современной информационной инфраструктуре.

Несовременность MARC

О недостатках MARC сказано много – это и плоская структура, и непонятность MARC-записей для поисковых машин и вне-библиотечного мира в целом, и целый ряд других ограничений, которые вызывали и вызывают критику. Справедливости ради стоит отметить, что часть указанных проблем решена в формате UNIMARC и форматах, на нем основанных (в частности, проблема «плоскости» записей – в UNIMARC широко используются связи данных как между записями, так и внутри записи). Другая часть проблем относится не к самому MARC-формату, а к ISO2709, и частично решается с помощью ISO 25577 MarcXchange, который обеспечивает:

- гибкость и легкую адаптируемость к стандартам каталогизации за счет отсутствия фиксированного набора полей и определенного синтаксиса;
- отсутствие жестких ограничений на размер записи;
- возможность включения в запись текстов, имиджей, звуковых файлов и т.д.;

- распространение вне библиотечного сообщества.

Дополнительные возможности связи данных реализуются в ISO 25577 за счет механизма встроенных полей UNIMARC, который был введен в стандарт по предложению Национальной службы RUSMARC.

Но в целом – да, вечного ничего нет, и в общем случае новые технологии при правильном их использовании обеспечивают больше возможностей для того, чтобы открыть библиографические данные для широкого интернет-сообщества за пределами библиотечных каталогов. Однако разработка новых инструментов – форматов, моделей, онтологий и т.п., – всегда требует значительных организационных, финансовых и интеллектуальных усилий. Необходима соответствующая административная поддержка, кардинальная перестройка системы обучения каталогизаторов / специалистов по метаданным, изменение технологической цепочки в библиотеках, разработка программного обеспечения для поддержки новых инструментов. Прежде чем говорить о полном отказе от MARC-формата, мы должны убедиться, что то, что предлагается ему на замену, имеет достаточно преимуществ, чтобы перевесить стоимость и трудоемкость перехода на новые инструменты и потенциально – миграции миллиардов записей из MARC. Располагаем ли мы достаточными ресурсами для такого перехода? Можем ли мы достичь цели другими средствами? Не ответив на эти вопросы, мы рискуем повторить наше любимое «весь мир насилья мы разрушим до основания, а затем мы наш, мы новый мир построим...». Вот только новый мир построить не всегда хватает сил, средств и знаний. Яркий пример – судьба Сводного каталога библиотек России.

Задачи

Вспомним, для чего создавались MARC-форматы? Изначально – по сути для обеспечения таких технологических процессов, как подготовка печатных каталогов и печать каталожных карточек. Отголоски этого до сих пор можно видеть в некоторых MARC-форматах, несмотря на их колоссальное развитие. Со временем задачи и возможности формата расширились – это и основа АБИС (а соответственно – и влияние на определение технологических процессов), и обмен метаданными, и обеспечение контролируемых точек доступа для эффективного поиска, и обеспечение иерархических и горизонтальных связей между библиографическими сущностями, и обеспечение работы с фондами (холдинговый формат). Появление формата для авторитетных данных означало еще больший отход от плоской структуры за счет возможности установления связей между авторитетными и библиографическими данными.

Если мы говорим о необходимости новых инструментов, которые должны обеспечить видимость библиотечных каталогов в интернете, мы должны четко определить, какие задачи должны решать эти инструменты, и насколько лучше они будут работать, чем MARC.

BIBFRAME

BIBFRAME с самого начала позиционировался как полная замена MARC-стандартов, новый способ представления библиографических данных и обмена ими, и шире – использование принципов связанных данных для того, чтобы сделать библиографические данные более полезными как внутри библиотечного сообщества, так и вне его.

Первая часть (представление библиографических данных и обмен) объясняет разработку детальных таблиц соответствия между MARC21 и BIBFRAME на уровне полей, подполей и даже отдельных позиций кодированных полей. Предполагается, что накопленные массивы MARC-записей должны быть преобразованы в ресурсы BIBFRAME. При этом провозглашается отход от каталогизации, ориентированной на воплощение, и переход от создания записей к созданию данных.

Соответственно онтология BIBFRAME также разрабатывается с высоким уровнем детальности, в расчете на то, что это поможет обеспечить и представление данных, и их обнаружение.

Сейчас BIBFRAME доведен до стадии практической реализации – разработаны не только модель и онтология, но и целый ряд инструментов для конвертирования MARC-форматов, ведется работа над редактором BIBFRAME. В той или иной мере BIBFRAME имплементирован в ряде библиотек различного уровня, другие присматриваются к нему и следят за развитием проекта. Однако многозадачность приводит к сложности реализации, некоторые решения представляются спорными.

В частности, создается собственный словарь BIBFRAME, в котором большинство терминов не привязывается к стандартным словарям, что отходит от рекомендаций W3C (см. «Best Practices for Publishing Linked Data» <https://www.w3.org/TR/ld-bp/>) – не изобретать новые термины, а дополнять существующие словари. Понятно, что для собственного спокойствия лучше держать свой словарь под контролем. И такой вариант допустим, но при условии, что словарь представляется по единым правилам, описанным в рекомендациях W3C, с установлением необходимых связей. В противном случае мы не сможем говорить о глобальной среде связанных данных.

Могу предположить, что это решение – одна из причин, по которой BIBFRAME не используется за пределами библиотечного сообщества (а MARC-форматы часто упрекают именно за это).

Еще одна проблема, которая тоже связана с «универсальностью» BIBFRAME и довольно часто отмечается в публикациях – сложность и излишнее теоретизирование, отрыв от практики. Понятно, что каталогизаторы должны развиваться, их надо учить, но, как известно, какой бы функциональной модель не была, ее успех полностью зависит от ее практичности и принятия в текущей деятельности.

"Когда профессия отказывается от практики ради теории, когда она пытается быть чем-то, чем она не является, когда она игнорирует очевидные кризисы, сосредотачиваясь на искусственных, когда она барахтается в специальной терминологии, когда растрчивает человеческий интеллект и драгоценное время на строительство воздушных замков, эта профессия рискует устареть."

Edmunds J. BIBFRAME as Empty Vessel. URL: <https://docplayer.net/51075426-Bibframe-as-empty-vessel-jeff-edmunds-23-february-2017.html>.

Schema.org

Еще один инструмент – Schema.org. Это стандартная схема, принятая OCLC (и многими другими организациями за пределами библиотечного сообщества), которая позволяет определять сущности и атрибуты способом, понятным для Интернета.

Schema.org – совместная инициатива ведущих интернет-компаний Bing, Google, Yahoo! и Yandex по разработке стандарта семантической разметки данных в сети. Цель создания Schema.org – помощь веб-разработчикам в создании качественных метаданных для повышения качества поиска. Schema.org может встраиваться в стандартный HTML-код, и метаданные, описанные с использованием Schema.org, могут быть напрямую проанализированы поисковыми роботами, помогая им лучше понимать содержимое веб-ресурсов. В настоящее время эта схема де-факто является стандартом структурированных данных в вебе.

Schema.org не предполагалась как глобальная онтология и тем более – как замена библиотечным стандартам; она обеспечивает широкое описание и не предполагает представления основных сущностей FRBR – «Произведение», «Выражение», «Воплощение» и «Физическая единица», в ней не проводится четкого различия между контентом и носителем, определено очень мало отношений между произведениями, и все это по простой причине – это не входит в круг задач Schema.org. Тем не менее,

Schema.org определяет достаточно согласованную модель с классами и свойствами, которые могут использоваться для простых описаний библиографических ресурсов, включая название, имя автора, место, организацию, издателя, дату авторского права, книгу, ISBN и т.д.

На базе Schema.org может разрабатываться более подробная онтология в той или иной предметной области – в частности, разработано библиографическое расширение.

Перспективным представляется возможное согласование библиотечных стандартов (в том числе и BIBFRAME) с Schema.org, аналогично модели представления произведений OCLC (описание модели см., например, Godby C.J., Wang S., Mixter J.K. Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description, 2015. – URL: <https://www.oclc.org/research/publications/2015/oclcresearch-library-linked-data-in-the-cloud.html>). В основе лежит подробное описание ресурса в библиотечном стандарте (MARC, BIBFRAME и т.д.). Ключевые понятия, которые релевантны для использования в Сети и необходимы для нахождения ресурса, преобразуются в RDF, используя Schema.org и расширение BiblioGraph.

Использование Schema.org обеспечивает ту самую открытость библиотечных данных для вне-библиотечного мира, к которой мы стремимся. Дженифер Зайно в статье «The Future of Libraries, Linked Data and Schema.Org Extensions» (<https://www.dataversity.net/the-future-of-libraries-linked-data-and-schema-org-extensions/>) цитирует Ричарда Уоллиса, председателя Community Group W3C по разработке библиографического расширения Schema.org: «Вы никогда не будете использовать [S]схему [.org] в качестве словаря для работы с ней библиотеки. Она не станет достаточно глубокой и богатой, чтобы охватить все тонкости данных MARC». И далее Дж. Зайно продолжает: «Но то, что она будет делать, – это помощь поисковой машине – размечать библиографические данные таким образом, чтобы поисковые приложения могли их понять и использовать соответствующим образом – и это то, что средний пользователь, вероятно, оценит больше.»

Британская национальная библиография

При разработке модели для представления в RDF Британской национальной библиографии (BNB) использованы стандартные схемы и существующие RDF-словари: Bibliographic Ontology, Dublin Core, FOAF, ISBD, Org, OWL, RDA, RDF Schema, Schema.org, SKOS, WGS84 Geo Positioning, а также British Library Terms – небольшой словарь, дополняющий существующие библиографические схемы. Метаданные

дополнены ссылками к другим наборам данных: GeoNames, ISNI, LCSH, перечням кодов MARC, VIAF, Wikidata и др. Использование существующих словарей (причем не только библиотечных, но и более широкого назначения) обеспечивает широкие возможности для взаимодействия как в библиотечном, так и внебиблиотечном интернете.

Интересно, что хотя Британская библиотека была в составе первой экспериментальной группы BIBFRAME, переход на BIBFRAME в ее ближайшие планы не входит. Это не означает отсутствие интереса к проекту, но Британская библиотека, как и многие другие, выжидает, пока развитие BIBFRAME не перейдет в достаточно стабильную стадию и не наберет «критической массы» пользователей, что могло бы оправдать усилия и затраты в условиях ограниченных ресурсов.

Модель Национальной библиотеки Франции

Национальная библиотека Франции идет по своему пути – исходя из того, что предоставление данных в RDF является ключом к возможности поиска и повторного использования в семантической сети; но при этом проводя четкое различие между способом создания данных и тем, как они хранятся и распространяются. С учетом этого положения, а также принимая во внимание экономические факторы, в Национальной библиотеке Франции несколько лет назад было принято решение для каталогизации продолжать использовать существующий формат INTERMARC, доработав его до полной поддержки FRBR – IFLA LRM.

Для представления данных из записей INTERMARC в RDF и организации поиска создается своя онтология в соответствии с правилами W3C.

Результат реализации этой модели – Data.bnf.fr. Данные из различных источников в формате MARC или EAD поступают в информационный хаб, где контент, ссылки и сервисы, представленные в соответствии с сущностями FRBR – IFLA LRM, группируются и преобразуются в RDF на основе стандартных словарей (DC, SKOS, RDA и FOAF), а также обогащаются ссылками к другим ресурсам, либо внутренним для BnF (основной каталог, цифровая библиотека Gallica, и т.д.) либо внешним (Сводный каталог французских академических библиотек SUDOC, WorldCat, VIAF, Wikipedia и др.). Таким образом, Data.bnf.fr не означает замену существующих каталогов и других разрозненных хранилищ, он объединяет содержащуюся в них информацию и делает ее понятной для поисковых машин, и это при использовании MARC-формата.

Есть много других примеров, других схем представления библиографических данных в RDF. Разные схемы могут вполне бесконфликтно сосуществовать в едином

открытом облаке. Главное, чтобы они создавались и были представлены там по единым правилам, описанным в рекомендациях W3C.

Проект Национальной службы развития системы форматов RUSMARC

Лет 5 назад Национальная служба развития системы форматов RUSMARC начинала работы по разработке подходов к представлению библиотечного каталога в семантическом вебе на базе национального формата RUSMARC. Очень коротко об основной идее: ядром системы являются RDF-записи авторитетного файла произведений (в терминах FRBR – IFLA LRM). Это ядро формируется на основе RUSMARC-данных и выполняет функцию информационного хаба. Помимо заглавия произведения, каждая такая запись хранит ссылки к АФ предметных рубрик, имен лиц и организаций, необходимую описательную информацию, а также корректно сформированный запрос к HTML-интерфейсу библиотечного каталога для выхода на список воплощений, имеющихся в библиотеке, и к библиотечным сервисам, доступным через стандартный интерфейс библиотечного каталога. Рабочую модель см. <https://lod-cloud.net/dataset/nlr-dataset>.

К сожалению, по ряду причин эта работа не была завершена, но, надеюсь, к ней можно будет вернуться в более практической плоскости.

Итог

Возвращаясь к нашей основной теме – судьбе MARC-форматов.

Разработка новых инструментов – форматов, моделей, онтологий и т.п. требует значительных организационных, финансовых и интеллектуальных усилий. Это под силу только крупным библиотекам или консорциумам, остальным приходится либо примыкать к проектам, которые реализуются ведущими игроками на информационном библиотечном поле, либо довольствоваться ролью зрителя и принимать уже готовый результат. В свое время стандарт USMARC (MARC21) самоутвердился в качестве базового стандарта структуры библиографических записей. Сейчас аналогичная ситуация складывается с BIBFRAME, несмотря на вопросы, о которых я говорила. Определенная логика в том, что BIBFRAME принимается другими библиотеками, есть. Помимо проработанности модели, это и заслуженный авторитет Библиотеки Конгресса, которая традиционно является одним из лидеров в библиотечном сообществе, и стремление не выпасть из передовой обоймы информационно-библиотечного сектора, и, не в последнюю очередь, экономические факторы. Что лучше – принять готовый инструмент

и использовать всеми или большинством членов сообщества, несмотря на его возможные недостатки, или пусть медленно, набивая собственные шишки, но идти по собственному пути, который представляется более правильным, но по которому есть риск не дойти до счастливого конца?

Еще один момент, о котором, к сожалению, нельзя не сказать, – колоссальная разница между ситуацией в России и за рубежом. Зарубежные библиотеки активно развиваются с точки зрения их видимости вне библиотечного интернет-сообщества (кто-то имплементировал BIBFRAME в той или иной степени (прежде всего здесь стоит сказать о НБ Швеции), кто-то, как Британская библиотека и НБ Франции – создает свою модель, кто-то, как НБ Финляндии, конвертирует MARC-данные в BIBFRAME, а затем извлекает данные в Schema.org для их представления в Сети; свой подход используется другими библиотеками, которые, тем не менее, присматриваются к BIBFRAME и следят за его развитием. Мы же только подходим к решению этих проблем, хотя понимание их важности сложилось не вчера.

Не будем говорить о причинах, но мы уже отстаем лет на 10 минимум. В настоящее время, к сожалению, у нас:

- нет согласованного понимания, как и куда двигаться;
- нет административной поддержки;
- нет системы обучения специалистов необходимой квалификации;
- нет программного обеспечения, которое бы поддерживало новый формат;
- нет гарантии сохранения информации при потенциальном переходе от RUSMARC на другие инструменты.

Пока не будут решены эти проблемы, говорить об отказе от MARC бессмысленно.

Умрет MARC или нет, зависит от того, какая схема связанных данных будет принята библиотечным сообществом. Как показывает практика, для полноценной реализации преимуществ связанных данных совсем необязательно, чтобы он умирал. И даже наоборот. В случае сохранения MARC-данных в качестве основы, их преобразования в RDF на основе стандартных словарей и обогащения связями с другими ресурсами, внутренними и внешними, кардинального изменения технологических процессов каталогизации и используемых систем не потребуется, но задача обеспечить видимость каталогов для вне-библиотечного интернет-сообщества и повысить практическую пользу библиотечных метаданных будет выполнена. При этом RUSMARC и другие форматы, основанные на UNIMARC, имеют большое преимущество за счет реализации механизма встроенных полей и широкого применения связи данных как

между записями, так и внутри записи. С точки зрения самой идеи Семантического веба и экономической выгоды лучше, если бы MARC продолжал жить, развиваться и использоваться в новых условиях.