

**Представление библиотечных данных в LOD:
возможности и перспективы формата RUSMARC**

Библиотечные каталоги остаются в большой степени изолированными от остального информационного поля и недоступными для поисковых машин в интернете. Решить эту проблему позволяет технология связанных данных, и за рубежом работы, направленные на представление библиотечных каталогов в виде связанных данных, ведутся практически с начала 21 века. Число опубликованных наборов библиотечных связанных данных растет, число библиотек, занимающихся проблемой связанных данных, тоже. Российские библиотеки долгое время оставались в стороне, но, кажется, эта ситуация наконец начинает меняться. Создание Рабочей группы для решения задач по разработке национальной модели библиотечных связанных данных по инициативе Секции по автоматизации, форматам и каталогизации и Секции «Электронные ресурсы и информационно-библиотечное обслуживание» РБА говорит о том, что российское библиотечное сообщество приступает к поиску путей, которые могут обеспечить интеграцию библиотечных данных в универсальное информационное пространство.

подавляющая часть библиотечных данных в России представлена в MARC-формате, как правило, это – национальный формат RUSMARC (за исключением ряда библиотек). Что нужно сделать, чтобы эти данные можно было использовать в среде связанных данных, и каковы перспективы формата RUSMARC в связи с переходом к новой технологии?

Связанные данные. В самом общем значении, как, собственно, и видно из названия, связанные данные – это коллекция взаимосвязанных структурированных наборов данных во Всемирной паутине. Когда говорят о связанных данных, также имеют в виду методы публикации связанных между

собой структурированных данных, позволяющие распространять информацию в машиночитаемом виде и обеспечивающие возможность поиска и работы с данными из разных источников.

По определению, *Связанные данные* – это наборы структурированных данных, опубликованные в RDF с использованием унифицированных идентификаторов ресурсов URI для идентификации сущностей и установления связей между отдельными сущностями.

Основные принципы связанных данных сформулировал создатель Всемирной паутины Тим Бернерс-Ли в 2006 году:

1. URI применяются в качестве имен объектов (при этом объекты – это не только библиографические записи или информационные ресурсы / документы, но и объекты реального мира, абстрактные понятия и т. д.).
2. URI представляются в формате протокола HTTP, чтобы можно было обращаться к объектам по этим URI.
3. Для описания информации и формулирования запросов к ней используются открытые стандарты – прежде всего RDF (для описания) и SPARQL (для запросов).
4. Обязательным является использование ссылок на другие URI, позволяющих найти дополнительную информацию.

Понятие «*открытые данные*» связано с возможностью доступа к данным, их использования и перепубликации пользователями в своих целях, без каких-либо ограничений в виде авторских прав, патентов и других механизмов контроля. Такая возможность реализуется с помощью свободных лицензий, таких как лицензии Creative Commons.

Когда говорят о *LOD, Linked Open Data* (открытые связанные данные), как правило, имеют в виду объединение связанных данных в глобальную распределенную открытую базу связанных данных, или в облако LOD Cloud.

Остановимся на тех аспектах процесса публикации связанных данных, которые так или иначе связаны с MARC-форматом.

Resource Description Framework. Технология связанных данных предполагает, что данные должны быть описаны в RDF (Resource Description Framework) – это среда описания связанных данных, которая позволяет технологии Semantic Web интерпретировать информацию, представленную в вебе.

Общая идея, на которой основана модель RDF, заключается в следующем: всё, что существует в мире (будь то физический предмет или абстрактное понятие), имеет определенные свойства, а любое свойство имеет конкретные значения. Значит, описать любую сущность можно с помощью элементарных выражений, которые называют эти свойства и их значения.

Основу модели RDF представляет трехчастное утверждение, или триплет, следующего вида:

Субъект – Предикат (или свойство) – Объект (значение свойства).

Триплет можно примерно соотнести с простым предложением вида *Подлежащее – Сказуемое – Дополнение*. Например, утверждение «*Книга написана Пушкиным*» в RDF-терминологии можно представить следующим образом: субъект – «*Книга*», предикат – «*написана*», объект – «*Пушкин*».

В отношении одного и того же субъекта могут быть составлены и другие выражения, определяющие другие его свойства.

Например, утверждение «*Книга «Полубрат», написанная Ларсом Сааби Кристенсеном, издана в 2007 году издательством «Иностранка»*. Тема книги отражена предметной рубрикой «*Семья в художественной литературе*»¹ можно представить в виде набора триплетов:

Ресурс – имеет название – «Полубрат»

Ресурс – имеет автора – Ларс Сааби Кристенсен

¹ В дальнейшем приводятся примеры из пилотного проекта по представлению каталога Российской национальной библиотеки в LOD, который был реализован Национальной службой развития системы форматов RUSMARC в 2014-2015 гг.

Ресурс – издан издательством – Иностранка

Ресурс – имеет дату издания – 2007

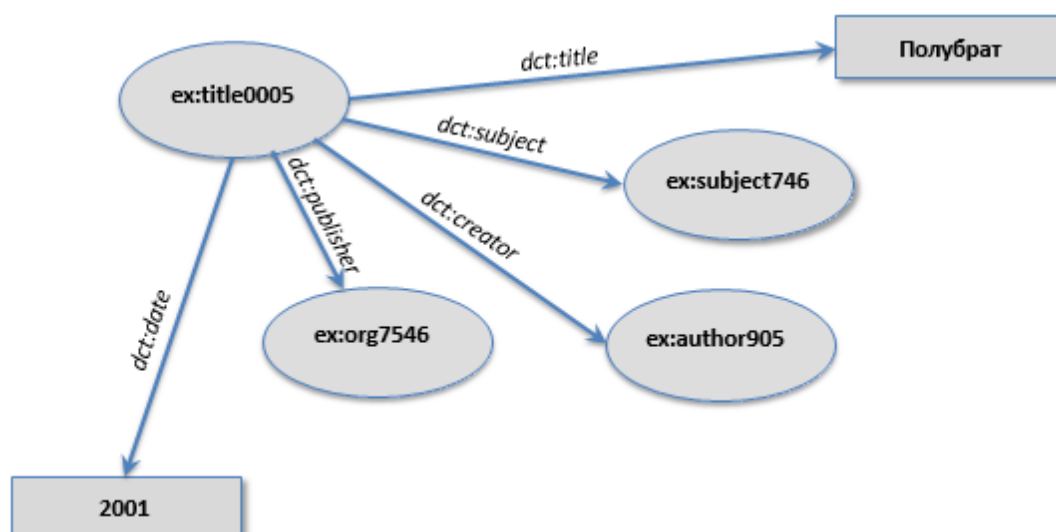
Ресурс – имеет тему – Семья в художественной литературе

Этот набор триплетов можно представить в виде графа, включающего субъект (ресурс) и 5 объектов, которые связаны с субъектом предикатами:

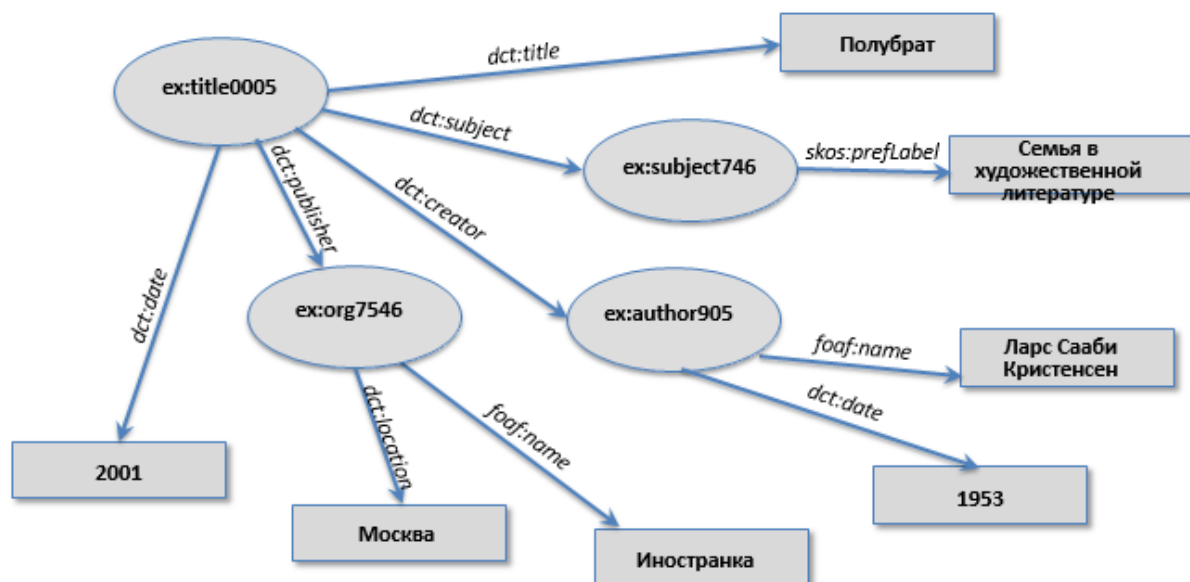


Каждой сущности (любому объекту реального мира или абстрактному понятию) вместо имени назначаются унифицированные идентификаторы ресурса (URI), чтобы по ним можно было обращаться к сущности. Предикаты определяются в соответствии с используемыми онтологиями.

Если ресурс имеет URI *ex:title0005*, автор, издательство и предметная рубрика также имеют соответствующие URI, а свойства, или предикаты, выражены в терминах Dublin Core, мы получаем:



Название книги и дата издания выражены литералами. С помощью операторов *skos:prefLabel* и *foaf:name* можно добавить литеральные значения для имени автора, названия издательства, предметной рубрики. Граф можно расширять, выстраивая отношения между сущностями.



Так, для автора может быть добавлен год рождения, для издательства – местонахождение – город Москва. Помимо того, что Москва является местонахождением издательства «Иностранка», она связана различными связями с другими сущностями (Москва - место рождения авторов, место проведения мероприятий, местонахождение памятников архитектуры, предмет художественных произведений и т.д.); предметная рубрика «Семья в художественной литературе» может быть присвоена другим произведениям; Ларс Сааби Кристенсен написал и другие произведения, которые изданы в различных издательствах как в оригинале, так и в переводе на другие языки, – т.е. за счет связей между сущностями можно осуществлять навигацию по графу от сущности к сущности.

То же граф, представленный в виде триплетов:

Субъект	Предикат	Объект
ex:title0005	dct:title	Полубрат
ex:title0005	dct:creator	ex:author905

ex:author905	foaf:name	Ларс Сааби Кристенсен
ex:author905	dct:date	1953
ex:title0005	dct:publisher	ex:org7546
ex:org7546	foaf:name	Иностранка
ex:org7546	dct:location	Москва
ex:title0005	dct:date	2007
ex:title0005	dct:subject	ex:subj734
ex:subj734	skos:prefLabel	Семья в художественной литературе

Здесь представлены 4 взаимосвязанные сущности – собственно произведение (а точнее, в данном случае – воплощение) с URI ex:title0006, автор с URI ex:author905, издательство с URI ex:org7546, и предметная рубрика с URI ex:subj734.

Преобразование MARC-записей в RDF заключается в выражении содержания записей (или части их содержания) в виде таких триплетов. Данные, выраженные в RDF, будут в дальнейшем выгружены в LOD.

Эта конечная задача предполагает решение нескольких подзадач, которые определяют, как будет выполняться процесс конвертирования записей из RUSMARC в RDF и как будут функционировать в конечном итоге связанные данные.

Модель данных. Прежде всего необходимо определить модель представления данных–набор формируемых сущностей и связи, которые будут установлены между ними, а также с внешними наборами данных. От этого зависит, какие данные из MARC-записей, в каком объеме подлежат конвертированию.

Например, в VIBFRAME предполагается, что записи конвертируются из MARC21 в RDF почти целиком, за исключением немногих элементов, которые разработчики посчитали неактуальными, а также элементов данных, которые релевантны только в MARC-окружении.

Другой вариант – представлять в RDF только те данные, которые требуются для обеспечения видимости библиотечных каталогов в глобальном интернете и установления необходимых связей. В таком варианте основная часть каталогов остается в исходном формате (например, MARC-формате, EAD и т.д.), не требуется полного изменения производственных процессов и правил описания, масштабного переучивания библиотекарей, замены программного обеспечения – по крайней мере, на переходном этапе. При этом задача сделать библиотечный каталог доступным и понятным для поисковых машин, открытым для небиблиотечного сектора интернета может быть решена вполне успешно.

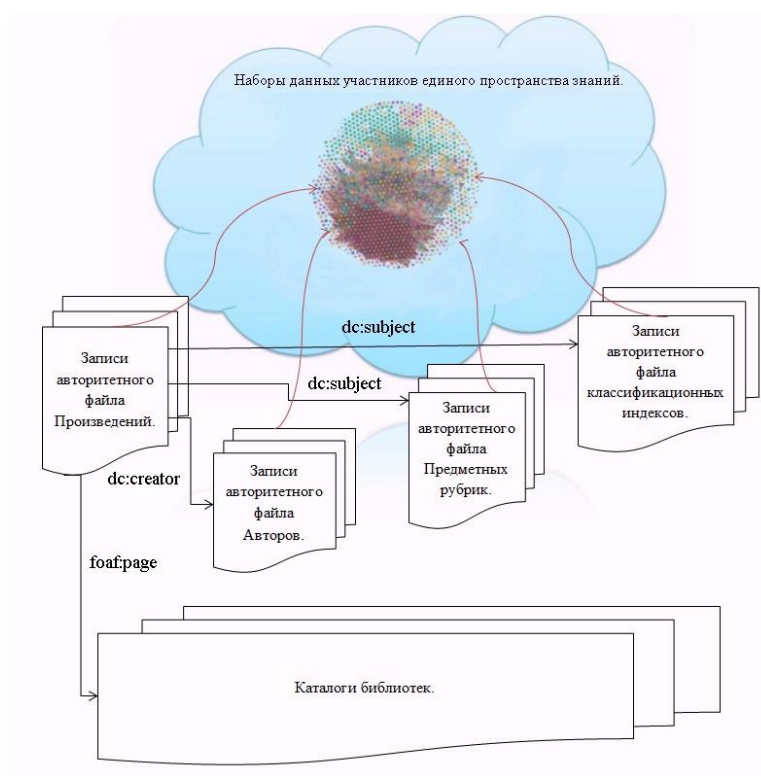
Такой вариант реализован, например, в Национальной библиотеке Франции. Веб-интерфейс Data.bnf.fr собирает описательные метаданные из разнородных источников (в основном представлены в формате MARC и EAD), трансформирует собранные метаданные в RDF-описания и представляет единую точку доступа к этим источникам. Метаданные в RDF могут быть проиндексированы поисковыми машинами и связаны с другими источниками – как внутри Национальной библиотеки Франции (генеральный каталог, Gallica и т.д.), так и вне ее (Сводный каталог французских научных библиотек Sudoc, WorldCat, VIAF, Wikipedia и др.). Итоговые данные публикуются в виде HTML-страниц и в виде RDF-дампа, доступного для загрузки.

В Национальной библиотеке Финляндии для публикации библиографических связанных данных выбрана модель Schema.org, представляющая общую информацию о ресурсах, а для конвертирования MARC-файлов в RDF используются инструменты BIBFRAME. После этого из полученных RDF-файлов извлекается информация о произведениях, эта информация преобразуется в Schema.org и публикуется в LOD.

В схеме, использованной в проекте Национальной службы RUSMARC², в LOD Cloud выгружаются RDF-файлы, которые формируются из исходных

² Скворцов В.В. Модель представления библиографических данных в системе связанных открытых данных (LOD). URL: <https://youtu.be/EgFy1HWVido>

библиографических записей для двух верхних уровней FRBR (файл *произведений* и *выражений*), а также авторитетные файлы *авторов* (индивидуальных и коллективных), *предметных рубрик* и *классификационных индексов*. *Воплощения* и *физические единицы* остаются в каталогах библиотек.



Ядро системы формируется на основе RUSMARC-данных и выполняет *функцию* информационного хаба. Каждый RDF-файл произведения, помимо заглавия, 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.

Такой подход соответствует логике модели FRBR (IFLA LRM) и в целом – логике развития процессов формирования и функционирования библиотечных каталогов, начиная с публикации модели FRBR и Международных принципов каталогизации. Контент (произведение и его различные версии: выражения) отделяется от физической формы (публикации, или воплощения и физической единицы); произведения и выражения

составляют основной уровень анализа библиотечных ресурсов и должны стать одной из основных точек входа в каталоги.

Выбор словарей. Следующий вопрос – выбор словарей и онтологий, используемых для описания библиотечных данных в рамках сформированной модели. Напрямую с форматом этот вопрос не связан, но поскольку библиотечные данные извлекаются из MARC-записей, или формируются на их основе, стоит сказать и об этом. При выборе словарей необходимо учитывать рекомендации W3C: использовать стандартные словари; если имеющихся словарей недостаточно – допускается создание собственного словаря, но он должен быть четко привязан к словарю стандартному.

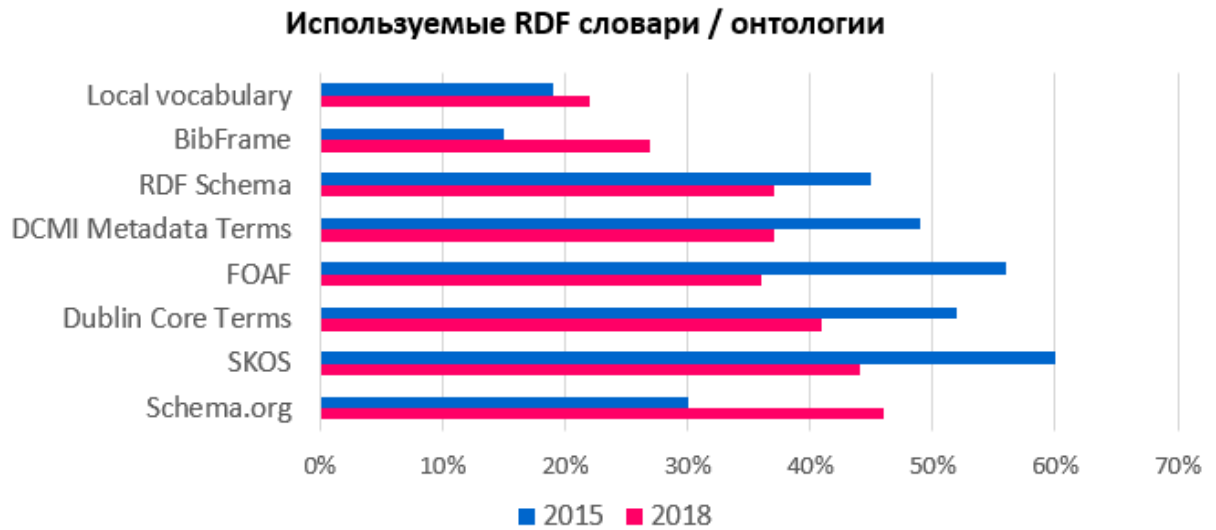
В 2014–2015 гг., и позднее в 2018 г. OCLC проводил опрос на тему связанных данных среди организаций, которые имплементировали или планируют хотя бы один проект, связанный с публикацией и / или использованием связанных данных. Среди участников представлены США, Испания, Великобритания, Нидерланды, Канада, Германия и др., – в основном научные и национальные библиотеки, исследовательские институты, библиотечные сети и консорциумы.

Один из вопросов был связан с выбором словарей, использованных для описания данных. Ниже на рисунке приведены 8 словарей, которые чаще всего упоминались в ответах респондентов. В 2015 г. наиболее часто использовались SKOS, FOAF и Dublin Core, в 2018 г. заметно увеличение использования Schema.org и BibFrame; чаще стали использоваться локальные словари.³

Кроме этих словарей, в ответах респондентов назывались словарь RDA, словарь модели данных Europeana, CIDOC CRM, FRBR, MADS, ISBD, MODS, ISNI и другие.

³ Karen Smith-Yoshimura. "Linked data implementations — who, what, why?" URL: <https://www.oclc.org/content/dam/research/events/2018/smith-yoshimura-linked-data-implementations-who-what-why-SWIB18.pptx>

OCLC: “International Linked Data Survey for Implementers” (2014-2015, 2018)



При этом выбор словарей для представления данных часто назывался в числе основных сложностей, с которыми приходится сталкиваться при публикации связанных данных.

Как правило, используется комбинация словарей. Полезный источник для выбора словарей – сайт LOV (<https://lov.linkeddata.es/dataset/lov/>), на котором собрана информация об основных словарях, используемых в LOD.

В проекте по представлению каталога РНБ в LOD использовались только стандартные словари RDFS, OWL, DC, SKOS, SKOS-XL, FOAF. Использование этих словарей позволяет сформировать основное описание, достаточное для идентификации произведения/выражения, предметных рубрик, имени автора. Список используемых словарей может быть расширен в дальнейшем, по мере необходимости.

Унифицированные идентификаторы ресурса. Один из основополагающих элементов концепции Связанных данных – это унифицированные идентификаторы ресурса (URI). URI должны быть функциональными – они служат именами объектов, по ним осуществляется обращение к объектам и по ним же устанавливаются связи.

Идентификаторы URI должны быть назначены каждому отдельному ресурсу (в широком смысле – не только информационному ресурсу, но и

любому объекту или даже абстрактному понятию, которые мы описываем). Для этого необходимо разработать правила, или шаблоны, формирования (назначения) URI своим ресурсам.

Общие рекомендации по использованию URI даны в документе «Cool URIs for the Semantic Web» (<https://www.w3.org/TR/cooluris/>). Сам термин «cool URI» ввел Тим Бернерс Ли в статье «Cool URIs don't change» (<https://www.w3.org/TR/cooluris/>). Главное условие – URI должен быть глобально уникальным и постоянным, поэтому при построении URI не следует включать в него информацию, которая может измениться со временем.

Например, в URI могут включаться:

- URL RDF-файла с описанием ресурса,
- указание типа ресурса,
- системный номер исходной авторитетной записи или библиографической записи, на основе которой строится RDF-описание (если такая запись существует).

Например, фрагмент авторитетной записи на имя автора (Ларс Сааби Кристенсен):

```
001 RU\NLR\AUTH\7735604
100 ##$a20060302arusy50#####ca
106 ##$a1
152 ##$aRCR
200 #1$aКристенсен$bЛ. С.$gЛарс Сааби$f1953-
300 0#$aNорвежский писатель, поэт, драматург; автор ряда
стихотворных сборников и пьес, нескольких сценариев
фильмов, рассказов и романов. Рецензент и редактор
выпускаемой издательством «Каппелен» антологии дебютов
«Signaler». Член Норвежской Академии языка и литературы.
```

В поле 001 записан идентификатор RU\NLR\AUTH\7735604, числовая часть которого может быть использована при построении URI: <http://www.rusmarc.ru/rdf/PersName/nlr7735604.rdf#person>.

Элементы URI: *rusmarc.ru/rdf/PersName* – это каталог, где хранятся файлы имен лиц, *nlr7735604.rdf* – имя файла, включающее префикс,

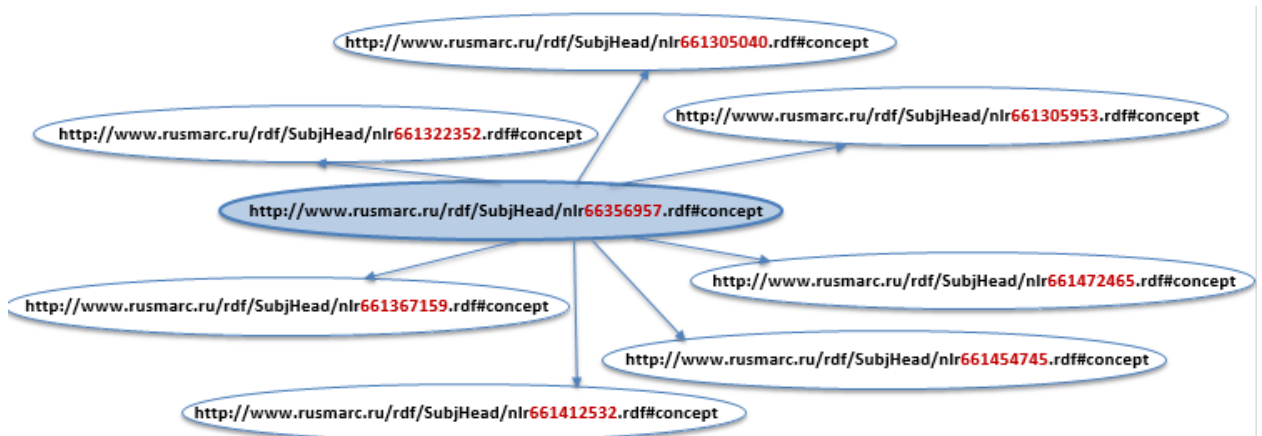
характеризующий систему, из которой получена исходная авторитетная запись (в данном случае – *nlr*), и системный номер этой авторитетной записи (из поля 001 авторитетной записи – 7735604), *#person* – добавление, обозначающее тип ресурса и образующее hash-URI лица.

Использование в составе URI системного номера авторитетной записи облегчает построение связей в RDF.

В качестве примера рассмотрим фрагмент авторитетной записи «Психолингвистика»:

```
001 RU\NLR\AUTH\66356957
152 ##$bnlr_sh
250 ##$aПсихолингвистика
550 ##$3RU\NLR\AUTH\661322352$5g$aМышление и речь$2nlr_sh
550 ##$3RU\NLR\AUTH\66356945$5h$aДетская речь$хРазвитие
    $хПсихологические исследования$2nlr_sh
550 ##$3RU\NLR\AUTH\661305040$5h$aЭтническая
    психолингвистика$2nlr_sh
550 ##$3RU\NLR\AUTH\661305953$5h$aЭтнопсихолингвистика$2nlr_sh
550 ##$3RU\NLR\AUTH\661367159$5h$aГендерная лингвистика$2nlr_sh
550 ##$3RU\NLR\AUTH\661412532$5h$aСуггестивная лингвистика
    $2nlr_sh
550 ##$3RU\NLR\AUTH\661472465$5h$aОнтолингвистика$2nlr_sh
550 ##$3RU\NLR\AUTH\661454745$5h$aЯзык$хОсвоение$2nlr_sh
```

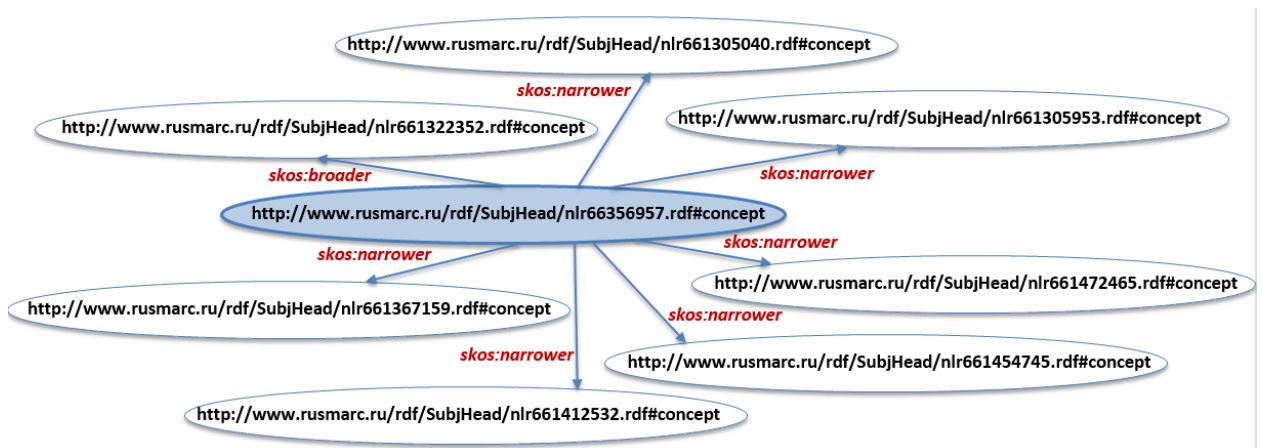
В полях 550 записаны связанные предметные рубрики, в подполе 550\$3 указаны идентификаторы авторитетных записей, включающих системный номер записи. Это позволяет вычислить URI связанных сущностей понятий и предметных рубрик и установить необходимые взаимосвязи.



Организация связей. Следующая задача – определить, какие взаимосвязи, и каким образом будут устанавливаться для нашего набора данных.

Связи (особенно внутренние, внутри публикуемого набора данных) могут строиться на основе тегов полей и других элементов данных MARC-записей (как авторитетных, так и библиографических).

Выше мы рассмотрели, как может для установления связей использоваться подполе \$3 в авторитетных записях. Тип связи, или предикат, в этом же примере может определяться на основе тега поля и значения кода в контрольном подполе \$5: при значении \$5/0 = g (более широкий термин) может использоваться предикат *skos:broader*, а при значении \$5/0 = h (более узкий термин) – соответственно *skos:narrower*.

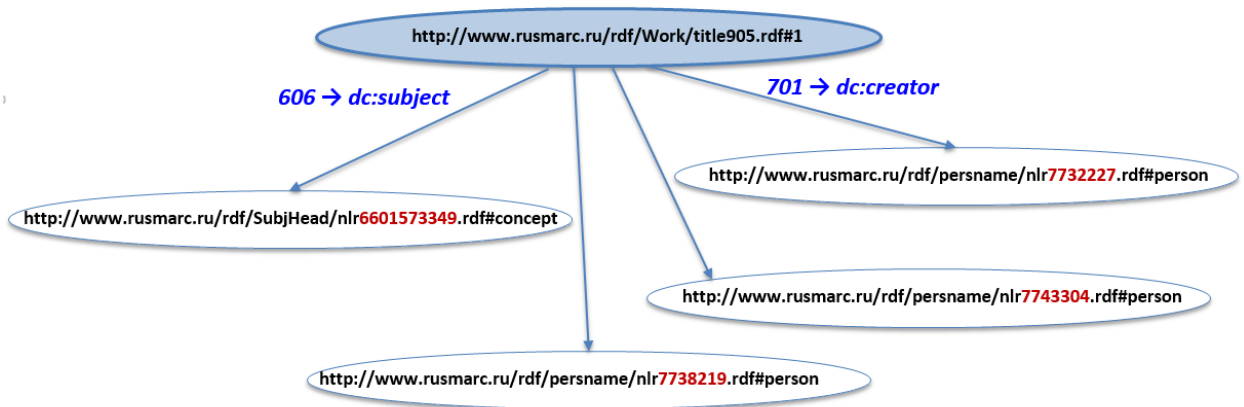


Еще один пример: при преобразовании библиографической записи в RDF тип связи определяется в соответствии с тегом поля. Тег поля 701 (соавторы) формирует связь *dc:creator*, поле 606 (предметные рубрики) – связь *dc:subject* (при использовании словаря Dublin Core).

```

200 1#$aСовременная библиотека в информационно-коммуникационной
    среде$fА. С. Тургаев, А. В. Соколов, И. И. Тихомирова и
    др.
606 1$$3RU\NLR\AUTH\6601573349aБиблиотековедение$уРоссийская
    Федерация
701 #1$3RU\NLR\AUTH\7732227$aТургаев$bА. С.$f1958-$gАлександр
    Сергеевич
701 #1$3RU\NLR\AUTH\7743304$aСоколов$bА. В.$f1934-$сд-р пед.
    наук$gАркадий Васильевич
  
```

701 #1\$3RU\NLR\AUTH\7738219\$aТихомирова\$bИ. И.\$gИраида
Ивановна\$f1933-\$сканд. пед. наук



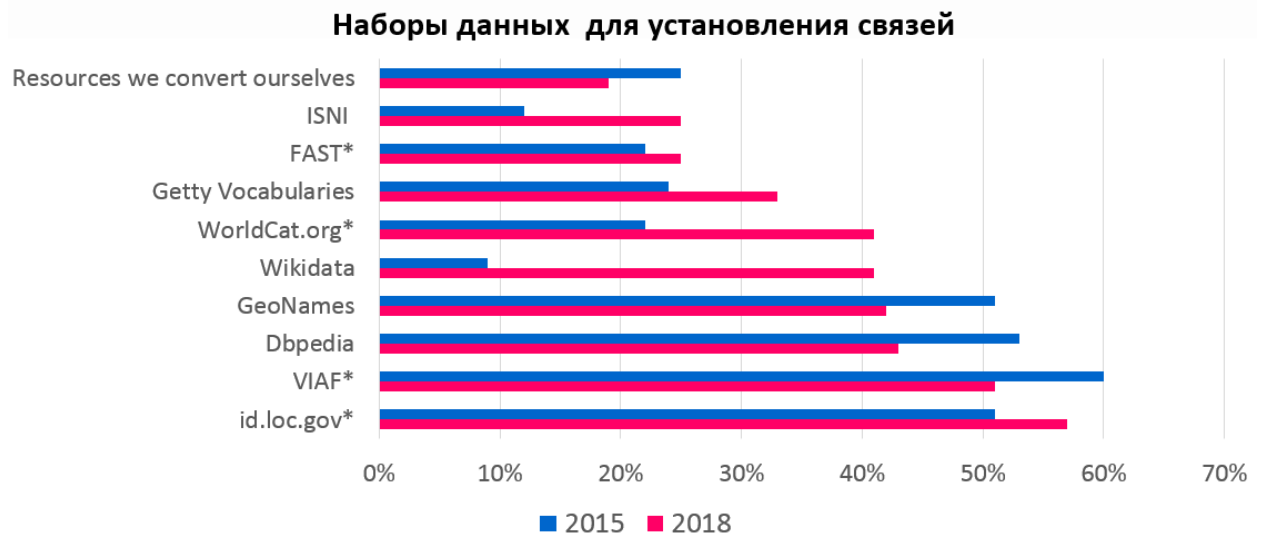
Для установления связей могут использоваться и другие элементы записи, в частности, стандартные номера и постоянные идентификаторы. Так, идентификатор ISNI, записанный в поле 010 авторитетной записи или в подполе \$o полей блока 7-- библиографической записи, позволяет построить алгоритм установления связи с соответствующим объектом (автор) в других каталогах при определенных условиях. Для предметных рубрик построение внешних связей сложнее, как, собственно, везде, где дело касается семантики и естественного языка. В свое время в РНБ проводился пилотный проект по установлению соответствия между предметными рубриками РНБ и LCSH (предметными рубриками Библиотеки Конгресса) на основе анализа частоты совместной встречаемости рубрик в библиографических записях. Если бы этот проект был полностью реализован, в авторитетных записях в полях 5-- блока были бы записаны связанные предметные рубрики LCSH с идентификатором записи и кодом системы, и это позволило бы хотя бы частично автоматизировать установление связей с набором LCSH.

Очевидно, что чем больше связей будет установлено как внутри публикуемого набора данных, так и внешних, тем проще будет переходить от узла к узлу, и тем более открыты будут наши данные для поиска. В настоящее время существует уже достаточно большой выбор наборов данных в RDF, которые можно использовать для установления связей. Часть из них

охватывает широкую тематику, другие сконцентрированы на отдельной категории данных или отдельной теме.

По данным опросов OCLC, которые упоминались выше, чаще всего для установления связей используются: каталог Библиотеки Конгресса (id.loc.gov), VIAF, Dbpedia, Geonames, Wikidata, каталог OCLC (Worldcat.org), словари Getty, FAST, ISNI и собственные ресурсы⁴.

**OCLC: “International Linked Data Survey
for Implementers” (2014-2015, 2018)**



Также в ответах респондентов упоминались: Europeana, Deutsche National Bibliothek Linked Data Services, Lexvo, WorldCat.org Works, data.bnf.fr, ORCID (Open Researcher and Contributor ID), DPLA (Digital Public Library of America) и Hispana.

Примерно треть респондентов по данным опросов OCLC назвали установление взаимосвязей в числе основных трудностей при подготовке публикации связанных данных.

Таблицы соответствия. На основе построенной модели представления данных в RDF, с учетом выбранных онтологий, шаблона построения URI, правил установления связей (той их части, которые могут быть установлены автоматически) строится таблица соответствия элементов данных RUSMARC

⁴ Karen Smith-Yoshimura. "Linked data implementations — who, what, why?" URL: <https://www.oclc.org/content/dam/research/events/2018/smith-yoshimura-linked-data-implementations-who-what-why-SWIB18.pptx>

и RDF, на ее основе в дальнейшем разрабатывается алгоритм и программа конвертирования. Высокая детализация данных в формате RUSMARC позволяет сделать таблицы соответствия достаточно подробными. Данные извлекаются из определенных полей, либо формируются на основе значений полей или подполей. Предикаты определяются, как правило, на основе тегов полей, идентификаторов подполей и значений кодированных данных, в рамках выбранных онтологий.

Ниже на рисунке приведен фрагмент таблицы соответствия между RUSMARC и RDF, построенной для преобразования авторитетных записей предметных рубрик.

RDF-запись на тематическую предметную рубрику создается в зависимости от сочетания кодов в маркере записи и поле 100.

Для каждого поля определено, информация из каких подполей берется при формировании RDF-записи и какой предикат какой онтологии при этом используется. В данном случае используется онтология SKOS. Заголовок и подзаголовки предметной рубрики записываются в *skos:prefLabel*, варианты формы – *skos:altLabel*. Для связанных предметных рубрик из полей 550 формируется связь с помощью предикатов *skos:broader* и *skos:narrower*, в зависимости от значения кода в подполе \$5.

Элемент данных	RUSMARC	RM/RDF	Примечания	Пример
Тип записи	Маркер, поз.симв.6 = x	RDF-запись формируется при указанных значениях	Авторитетная запись	
Тип объекта описания	Маркер, поз.симв.9 = j		Наименование темы	
Статус точки доступа	100/\$a/8 = a		Установленная	
Идентификатор АЗ	001	<pre><rdf:Description rdf:about="http://rusmarc.ru/ rdf/SubjHead/nlr____.rdf#concept"> <rdf:type rdf:resource= "http://www.w3.org/2004/02 /skos/core #Concept"/> </rdf:Description></pre>	Цифровая часть идентификатора записи в поле 001 RUSMARC записи-донора используется в качестве элемента URI RDF-записи	<pre><rdf:Description rdf:about="http://rusmarc.ru/rdf/SubjHead/ nlr6678578.rdf#concept"></pre>
Наименование темы - принятая форма	250\$a, 250\$j, 250\$x, 250\$y, 250\$z	skos:prefLabel	Подполя включаются в том порядке, в котором приведены в записи-доноре RUSMARC, и разделяются символами «-».	<pre><skos:prefLabel xml:lang="ru">Роман норвежский--20 в.</skos:prefLabel></pre>
Публикуемое примечание	300\$a, 305\$a\$b, 310\$a\$b, 320\$a, 330\$a	skos:scopeNote	Подполя включаются в том порядке, в котором приведены в записи-доноре RUSMARC, и разделяются символами пробела.	<pre><skos:scopeNote>См. также под названием отдельных предметов с подзаголовком "Энциклопедии", например: Техника - Энциклопедии</skos:scopeNote></pre>
Вариантные формы	450\$a, 450\$j, 450\$x, 450\$y, 450\$z	skos:altLabel	Подполя включаются в том порядке, в котором приведены в записи-доноре RUSMARC, и разделяются символами пробела.	<pre><skos:altLabel>Ирако-Кувейтский конфликт--1990 - 1991</skos:altLabel></pre>

Связанные термины – более широкие	550\$3 Если 550\$5/0 = g	skos:broader	Цифровая часть идентификатора записи в подполе 550\$3 используется в качестве элемента URI RDF-записи	<skos:broader rdf:resource="http://www.rusmarc.ru/rdf/SubjectHead/nlr66151610.rdf#concept"/>
Связанные термины – более узкие	550\$3 Если 550\$5/0 = h	skos:narrower	Цифровая часть идентификатора записи в подполе 550\$3 используется в качестве элемента URI RDF-записи	<skos:narrower rdf:resource="http://www.rusmarc.ru/rdf/SubjectHead/nlr661563410.rdf#concept"/>
Связанные тематические термины	550\$3 Если 550\$5/0 != g и h или 550\$5 отсутствует	skos:related	Цифровая часть идентификатора записи в подполе 550\$3 используется в качестве элемента URI RDF-записи	<skos:related rdf:resource="http://www.rusmarc.ru/rdf/SubjectHead/nlr661563993.rdf#concept"/>
Источник, в котором выявлена информация о точке доступа	810\$a\$b	skos:editorialNote	Подполя включаются в том порядке, в котором приведены в записи-доноре RUSMARC, предваряются литералом «Источник:» и разделяются точкой.	<skos:editorialNote>Ист.: Мат.энциклопедия, т.2., 1979. Интеграл по траекториям, континуальный интеграл, функциональный интеграл, - интеграл, областью интегрирования которого служит таким образом то или иное функциональное пространство</skos:editorialNote>
Общее примечание каталогизатора	830\$a	skos:editorialNote		<skos:editorialNote>Термин уточнен в универсальном и отраслевом авторитетных источниках</skos:editorialNote>

Формирование RDF-записей. Последний этап – собственно конвертирование, в результате которого формируется массив RDF-записей. Часть из этих записей формируется из авторитетных записей на имя лица, тематические предметные рубрики. Описание произведений и выражений формируется из библиографических записей, имеющих в каталоге, с установлением связей на основании тегов полей RUSMARC и последующим добавлением дополнительных связей.

```

<!-- Work 1 in Russian -->
<rdf:RDF>
  <rdf:Description rdf:about="http://www.rusmarc.ru/RDF/Work/title1.rdf">
    <!-- Provenance metadata -->
    <dc:creator rdf:parseType="Resource">
      <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
      <rdflabel xml:lang="ru">
        Национальная Служба развития системы форматов RUSMARC
      </rdflabel>
      <foaf:homepage rdf:resource="http://www.rusmarc.ru/">
    </dc:creator>
    <dc:created rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2015-01-16</dc:created>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.rusmarc.ru/RDF/Work/title1.rdf#1">
    <!-- Book 1 in Russian -->
    <rdf:type rdf:resource="http://www.w3.org/2000/01/rdf-schema#Class"/>
    <dc:title>Полубрат</dc:title>
    <dc:type>http://purl.org/dc/dcmitype/Text</dc:type>
    <dc:creator rdf:resource="http://www.rusmarc.ru/RDF/PersName/nlr7735604.rdf#person"/>
    <!-- Христеисен -->
    <dc:language>ru</dc:language>
    <rdflabel>Произведение "Полубрат" на русском языке</rdflabel>
    <dc:subject rdf:resource="http://www.rusmarc.ru/RDF/SubjectHead/nlr661425274.rdf#concept"/>
    <!-- Литература норвежская 20 в. -->
    <dc:subject rdf:resource="http://www.rusmarc.ru/RDF/SubjectHead/nlr661006677.rdf#concept"/>
    <!-- Семья в художественной литературе -->
    <foaf:page rdf:resource="http://primo.nlr.ru/primo_library/libweb/action/dlSearch.do?institution=07NLR&vid=07NLR_VU1&tab=default_tab&indx=1&bulkSize=10&dym=true&highlight=true&displayField=title&query=any%2Ccontains%2C%D0%A%D1%80%D0%B8%D1%81%D1%82%D0%B5%D0%BD%D1%81%D0%B5%D0%BD+%D0%F%D0%BE%D0%BB%D1%83%D0%B1%D1%80%D0%B0%D1%82"/>
    <owl:sameAs rdf:resource="http://www.rusmarc.ru/work/title2.rdf#1"/>
  </rdf:Description>
  <rdf:Statement rdf:about="#triple12345">
    <rdf:subject rdf:resource="http://www.rusmarc.ru/RDF/Work/title1.rdf#1"/>
    <rdf:predicate rdf:resource="http://purl.org/dc/elements/1.1/description"/>
    <rdf:object>
      На страницах «Полубрата» уместилось полвека — с конца Второй мировой до рубежа тысячелетий. В центре сюжета двое сводных братьев, связанных странной, ожесточенной и болезненной любовью-ненавистью. Братьев окружают женщины, сильные, страстные и потому одинокие: мать, которая в день окончания войны стала жертвой насилия, бабушка и прабабушка-актриса немного кино, не сыгравшая ни одной роли. Это не просто семейная сага с захватывающим сюжетом, это еще и роман о лжи и самообмане, подчиняющих себе жизнь героев и разрушающих ее изнутри.
    </rdf:object>
    <dc:source rdf:resource="http://fb2lib.net.ru/book/82041"/>
  </rdf:Statement>
</rdf:RDF>

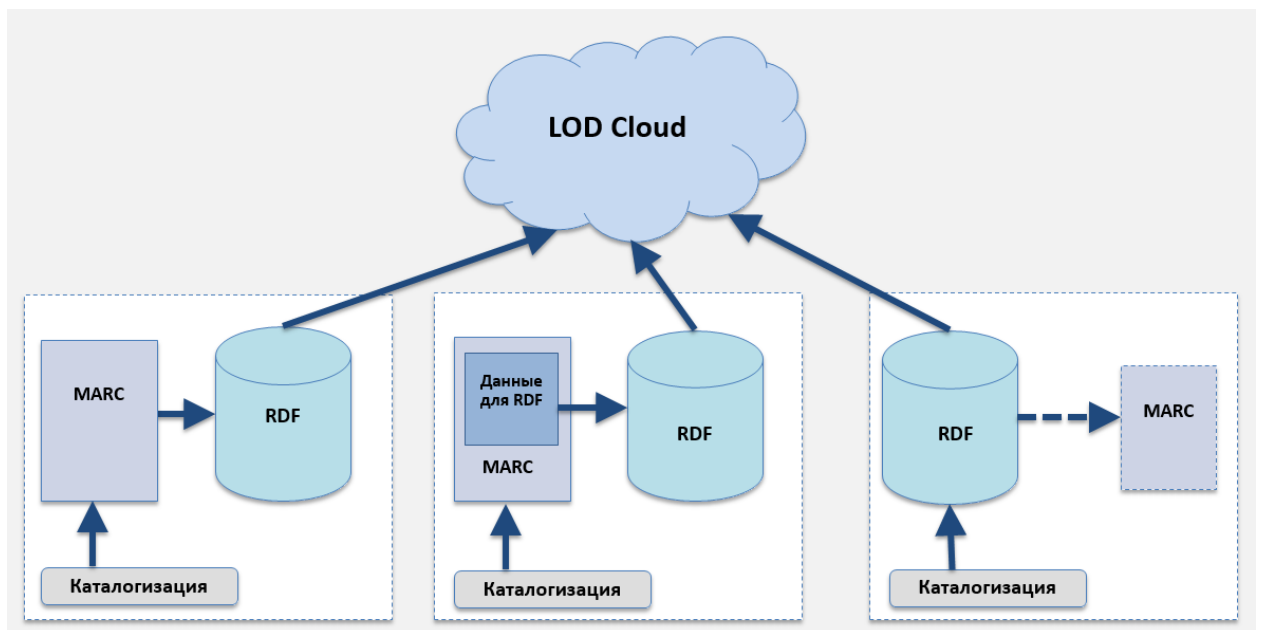
```

Напомним, что описание воплощений и физических единиц в этом случае остается в каталогах библиотек, которые по-прежнему ведутся в формате RUSMARC или другом исходном формате, по крайней мере в обозримом

будущем. Если такая модель будет принята, формат RUSMARC останется на ближайшее время с нами, и потребуются определенные усилия для его поддержки.

Даже формат MARC21, при том, что BIBFRAME активно продвигается как его замена, продолжает развиваться – достаточно посмотреть информацию о деятельности Консультативного комитета MAC (орган, ответственный за поддержку формата MARC21, в состав которого входят представители Библиотеки Конгресса, Национальной библиотеки и архивов Канады, Британской библиотеки, Немецкой национальной библиотеки).

До этого речь шла в основном о представлении в LOD MARC-данных из уже существующих каталогов. Что ждет при переходе к технологии связанных данных остальные технологические библиотечные процессы, прежде всего процесс каталогизации?



Один из вариантов – каталогизация по-прежнему осуществляется в MARC-формате, после этого данные конвертируются в RDF и передаются в LOD.

Возможен вариант такой схемы – трансформируются в RDF и передаются в LOD только те данные, которые необходимы для организации

доступа извне к библиотечным каталогам (и, естественно, обратно) и осуществления эффективного поиска. Это вариант, который реализован в описанной выше модели.

Большинство экспериментов со связанными данными не затрагивает процессы текущей каталогизации. Однако возможен вариант, при котором ведение каталога и текущая каталогизация осуществляются непосредственно со связанными данными.

Такая технология реализована в Публичной библиотеке Осло. Библиотечный каталог функционирует полностью в виде связанных данных с 2011 г.; в 2015 г. библиотека перешла на открытое программное обеспечение Koha, при этом было решено отказаться от каталогизации в MARC и каталогизировать непосредственно в RDF. Модули поиска и текущей каталогизации (Catalinker) ориентированы на FRBR и основаны на связанных данных RDF. Холдинговая информация, процессы книговыдачи и обслуживания пользователей поддерживаются модулем Koha. Библиографические метаданные в Koha представляют собой минимальный набор из MARC21 с RDA, которые генерируются из связанных данных и синхронизируются с ними. MARC-данные используются для генерации статистики, требуемой национальной библиотекой (статистика основана на фасетах формата NORMARC).

Как видите, о полном отказе от MARC-формата и в этом случае говорить не приходится. И это понятно – на формате завязаны технологические процессы, на формате построены АБИС, накоплены огромные информационные массивы.

Когда речь идет о преобразовании MARC-данных в RDF и использовании их в среде связанных данных, используется термин “legacy data” – унаследованные данные, т.е. информация, которая хранится в устаревших форматах или информационных системах. Как известно, наследство принимается целиком – включая и активы, и долги, и уже от

наследников зависит, как они распорядятся наследством – смогут ли избавиться от долгов и преумножить активы. При переходе к технологиям LOD мы получаем MARC-данные тоже со всеми достоинствами, которые можно и нужно использовать, и недостатками, которые нужно оставить в прошлом. Все возможности для этого у MARC-формата (в том числе и RUSMARCa) есть, а более предметно о перспективах формата можно будет говорить только после того, как будет определена национальная модель представления библиотечных связанных данных.