

ЖЛОБИНСКАЯ ОЛЬГА НИКОЛАЕВНА
Санкт-Петербург, Россия

ВОЗМОЖНОСТИ И ПЕРСПЕКТИВЫ ИСПОЛЬЗОВАНИЯ ТЕХНОЛОГИИ СВЯЗАННЫХ ДАННЫХ В ПРЕЗИДЕНТСКОЙ БИБЛИОТЕКЕ

Ключевые слова: библиотечный каталог, интегрированный каталог, связанные данные, архивные материалы, схемы метаданных, RUSMARC, IFLA LRM, «Документы в контекстах».

Keywords: library catalogue, integrated catalogue, linked data, archival materials, metadata schemas, RUSMARC, IFLA LRM, “Records-in-Contexts.”

Президентская библиотека в Санкт-Петербурге – общегосударственное электронное хранилище цифровых копий важнейших документов по истории, теории и практике российской государственности, русскому языку; основная часть фонда – это цифровые копии ресурсов из различных учреждений науки и культуры России и мира.

В настоящее время электронный фонд библиотеки насчитывает свыше 1 млн. единиц хранения. Помимо традиционных библиотечных материалов, в фонд поступают материалы из архивов и музеев. Доля библиотечных материалов в фонде Президентской библиотеки – 52%, архивных материалов – 46%; 2% – это материалы музейного хранения. Для раскрытия фонда ведется интегрированный электронный каталог, который обеспечивает описание, поиск и дисплейное представление различных типов и видов ресурсов.

Основные принципы функционирования электронного каталога Президентской библиотеки:

- единый формат представления данных в машиночитаемой форме – национальный формат RUSMARC;
- согласованная методика описания и индексирования, которая определяется источником поступления цифровой копии, а также типом хранения и описания документа-оригинала;
- использование единых авторитетных файлов для обеспечения контроля точек (имен лиц, наименований организаций, географических названий, предметов) для материалов библиотечного, архивного и музейного хранения;
- единый поисковый интерфейс, обеспечивающий поиск и просмотр разнородных ресурсов.

Построение и функционирование интегрированного каталога, объединяющего информацию о ресурсах различных областей культурного наследия, отвечает одной из наиболее актуальных задач в области сохранения культурного наследия – задаче создания единого информационного пространства библиотек, архивов и музеев.

Однако, по данным исследований от 15 до 20% пользователей начинают поиск не с библиотечных каталогов, а с Google или других поисковиков – такой путь поиска представляется пользователям и удобнее, и быстрее. Библиотечные каталоги, несмотря на все богатство заключенной в них информации, в силу ряда причин остаются в большой степени изолированными от остального информационного поля. Встает необходимость найти такие решения, которые помогли бы максимально открыть библиотечные данные для поиска в Интернете, обеспечить их интеграцию в универсальное информационное пространство. Это позволит библиотекам сохранить свою актуальность и место в современном мире.

Наиболее перспективным в этом отношении представляется использование концепции связанных данных – методов размещения структурированных данных в Сети и установления связей между ними таким образом, что данные могут быть не только найдены и поняты и человеком, и компьютером; при этом компьютер не просто сопоставляет строки текста, но и понимает его смысл. Более того, на основе формализованных связей компьютер может понимать отношения между объектами и объединять эту информацию, генерировать новые знания, в конечном счете обеспечивая возможность включения библиотечных данных в Глобальную систему знаний.

На национальном уровне задача объединения информационных ресурсов России в единое национальное цифровое культурное пространство на основе связанных данных зафиксировано в «Стратегии развития библиотечного дела в Российской Федерации на период до 2030 года» [Стратегия развития библиотечного дела ...: 2021].

В 2021 г. по инициативе Секции по автоматизации, форматам и каталогизации, Секции «Электронные ресурсы и информационно-библиотечное обслуживание» Российской библиотечной

ассоциации (РБА) и Межрегионального комитета по каталогизации была создана Рабочая группа по разработке и внедрению Национальной модели связанных библиографических данных для обеспечения интеграции всех библиографических данных на основе новых национальных стандартов в единый каталог с возможностью поиска и доступа к данным из различных приложений Semantic Web. В состав Рабочей группы вошли представители российских библиотек различного уровня, в том числе и представители Президентской библиотеки. Среди основных сложностей, с которыми сталкиваются члены Рабочей группы, можно назвать недостаток квалифицированных специалистов и в целом – низкую осведомленность библиотекарей в этой области, а также недостаточную заинтересованность со стороны структур, которые могли бы обеспечить организационную и финансовую поддержку.

Параллельно деятельности Рабочей группы РБА исследования в области связанных данных и Semantic Web ведутся и непосредственно в Президентской библиотеке.

Несколько лет назад Национальная служба развития системы форматов RUSMARC начинала работы по разработке подходов к представлению библиотечного каталога в среде связанных данных на базе национального формата RUSMARC и концептуальной модели библиографических данных FRBR (Functional Requirements for Bibliographic Records, Функциональные требования к библиографическим записям) [Functional requirements: 1998].

Подход, использованный в проекте Национальной службы RUSMARC [Скворцов : 2021], соответствует логике развития процессов формирования и функционирования библиотечных каталогов, начиная с публикации модели FRBR и Международных принципов каталогизации [Statement of International Cataloguing Principles: 2009]. Контент (в терминологии FRBR – 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.

Аналогичная схема может быть реализована и на материале электронного каталога и авторитетного файла Президентской библиотеки. При этом ряд вопросов, связанных со спецификой Президентской библиотеки, требуют отдельного рассмотрения.

1. Профиль комплектования Президентской библиотеки. Фонд Президентской библиотеки формируется преимущественно из электронных копий ресурсов из различных библиотек и других учреждений науки и культуры; все ресурсы представлены в фонде в единственном экземпляре. Если говорить в терминах модели FRBR (и ее современной ипостаси – IFLA LRM [IFLA Library Reference Model: 2017]), каждое произведение в фонде библиотеки и соответственно в каталоге представлено, как правило, одним воплощением одного выражения, что отличает каталог Президентской библиотеки от каталогов универсальных библиотек и тем более – от сводных каталогов, объединяющих сведения об экземплярах, хранящихся в различных библиотеках. Такая схема не противоречит FRBR, однако при этом еще больше возрастает роль внешних связей, т. е. связей информационных ресурсов Президентской библиотеки с ресурсами, представленными в каталогах и фондах других библиотек.

2. Авторитетный файл. Авторитетный файл Президентской библиотеки ведется с 2010 г., в настоящее время его объем составляет более 45 тысяч записей различных категорий, в том числе предметные рубрики, а также нормированные точки доступа на имена лиц и наименования организаций (в том числе фондообразователей) и географические названия. Технологические особенности формирования и использования авторитетного файла Президентской библиотеки обуславливают недостаточное количество зафиксированных внутренних связей; кроме того, полностью отсутствуют связи с другими словарями контролируемых терминов. Однако именно установление связей между различными наборами данных обеспечивает само существование связанных данных; установление связей – операция абсолютно необходимая, хотя, как показывает практика, очень трудоемкая и финансово затратная.

3. Архивные материалы. 46% материалов в фонде и электронном каталоге Президентской библиотеки составляют архивные материалы, для которых, в отличие от традиционных библиотечных ресурсов, характерна строгая иерархическая организация и контекстуализация – каждая архивная единица помещается в контекст единицы более крупной, и только в этом контексте она может быть адекватно оценена и воспринята.

Реализация иерархической организации архивных материалов потребовала особых решений в интегрированном каталоге и доработки машиночитаемого формата RUSMARC. В настоящее время эта задача решена. Архивные материалы представлены в каталоге Президентской библиотеки в точном соответствии с исходной структурой фондов, описей, дел, используемых в архиве-фондодержателе. При просмотре результатов поиска пользователю предоставлена возможность не только вывести на экран собственно цифровую копию, но также просмотреть найденные материалы в любой иерархии:

как сверху вниз, спускаясь от фонда к описям, затем к делам и к конкретным документам, так и снизу вверх, поднимаясь от конкретных документов, дел, описей к фонду в целом.

В процессе трансформации исходных MARC-записей в связанные данные сведения об иерархической структуре и историческом контексте создания материалов необходимо сохранить, наряду со сведениями непосредственно об архивной единице (фонде, описи, деле или документе) – фондодержателе, хронологических рамках, содержании материалов и т. д.

Описанная выше схема преобразования данных, разработанная Национальной службой RUSMARC, основана на концептуальной модели FRBR (IFLA LRM), которая была разработана специально для библиотечных данных и специфику архивных и музейных материалов не учитывает. Для архивных материалов представляет интерес концептуальная модель и онтология RiC (Records-in-Contexts, Документы в контекстах) [Records in Contexts – Conceptual Model: 2021; Records in Contexts – Ontology: 2021], разработанная Экспертной группой по архивному описанию (EGAD) Международного совета по архивам (ICA).

Трансформацию интегрированного каталога Президентской библиотеки предполагается осуществлять на базе модели данных, учитывающей и IFLA LRM, и RiC-CM. Для этого необходимо провести семантическое и структурное сопоставление двух моделей IFLA LRM и RiC-CM и определить оптимальный уровень детализации их согласования для интегрированного каталога. Например, в RiC-CM набор сущностей и связей обширнее, чем в IFLA LRM, но должен ли библиотечный каталог отражать все эти сущности и связи, определенные для архивных материалов, или достаточно оставить только основные? Если да – какие связи считать основными? Требуется ли точное соответствие между элементами данных, их атрибутами и связями между ними, или достаточно некоторой степени условности? Интегрированный библиотечный каталог не претендует на полное архивное описание, и определенная степень условности, по мнению автора, допустима, если она не ведет к снижению эффективности поиска. Но в целом эти вопросы требуют более детального анализа.

Тема связанных данных в российских библиотеках вызывает большой интерес, но пока, к сожалению, приходится признать, что этот интерес остается преимущественно в теоретической плоскости. Ряд причин уже назывались выше – нехватка квалифицированных специалистов, а также отсутствие инструментов, стандартов и вообще достаточной практики в области связанных данных и Semantic Web. Из реально действующих проектов можно назвать проект Российской государственной библиотеки по представлению Библиотечно-библиографической классификации в среде связанных данных для оптимизации поиска [Лавренова : 2021]. Еще один проект – Единое научное информационное пространство (ЕНИП) Российской академии наук. Если говорить о национальном архивном сообществе – автору не удалось найти информацию о каких-либо проектах с применением концепции связанных данных. В то же время реализация концепции связанных данных может обеспечить важные инструменты повышения функциональной совместимости цифровых данных культурного наследия. Президентская библиотека в силу специфики фонда, объединяющего информацию о ресурсах различных областей культурного наследия, призвана играть особую роль в решении задачи интеграции информационных ресурсов на современном технологическом уровне.

SUMMARY

The article discusses approaches to publishing library data of the Presidential Library in the LOD (Linked open data) environment as an integral part of a single digital knowledge space. The basic model of data is built taking into account the specifics of the Presidential Library catalogue: the integrated library catalogue includes library, archival and museum materials; when building the catalog, the national RUSMARC format is used. The proposed scheme for publishing linked data is based on the transformation of MARC records using the IFLA LRM and RiC-CM conceptual models as well as standard ontologies.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Functional requirements for bibliographic records : final report. München, Saur, 1998, VIII, 136 p. [Electronic Publication] URL: https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr_2008.pdf (Sept. 1997, as amended and corrected through February 2008), 28.11.2022.
2. IFLA Library Reference Model: A Conceptual Model for Bibliographic Information, 2017. [Electronic Publication] URL: https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf, 28.11.2022.

3. Records in Contexts – Conceptual Model, 2021. 121 p. [Electronic Publication] URL: https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf, 28.11.2022.
4. Records in Contexts – Ontology, 2021. [Electronic Publication] URL: https://www.ica.org/standards/RiC/RiC-O_v0-2.html, 28.11.2022.
5. Statement of International Cataloguing Principles (ICP). 2009. [Electronic Publication] URL: https://cdn.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/icp/icp_2009-en.pdf, 28.11.2022.
6. Лавренова О. А. Опыт представления ББК в LOD // Теория и практика каталогизации и поиска библиотечных ресурсов: электронный журнал, 2017. Эл. рес.: <http://www.nilc.ru/text/NMLBD/NMLBD6.pdf>, 28.11.2022
7. Скворцов В. В. Модель представления библиографических данных в системе связанных открытых данных (LOD). 2021. Эл. рес.: <https://youtu.be/EgFy1HWVido>, 28.11.2022.
8. Стратегия развития библиотечного дела в Российской Федерации на период до 2030 года. Эл. рес.: <http://static.government.ru/media/files/NFWPpXpAAAEbPW60HiZiDvdZZ8AcSNuu.pdf>, 28.11.2022.