



Немецкая национальная библиография как открытые связанные данные: использование и возможности

Jürgen Kett

Sarah Beyer

Mathias Manecke

Yvonne Jahns and

Lars G. Svensson

Немецкая национальная библиотека
Франкфурт-на-Майне, Германия

*Перевод с сокращениями А.В. Теплицкой
(Российская государственная библиотека, Москва)*

Session:

215 — What is a national bibliography today and what are its potential uses? — Bibliography

Сессия: 215 – *Что такое национальная библиография сегодня, и какая от нее польза – Библиография*

Реферат:

В докладе обсуждаются требования, предъявляемые к национальной библиографии в 21-м веке. Учитывая современные информационные изменения, национальной библиографии, построенной на традиционных критериях полноты и надежности данных, оперативности, репрезентативности (referentiability)¹, стабильности, использующей авторитетные данные, необходимо стать составной частью всемирной паутины. Это можно сделать на основе технологий связанных данных, и эти данные должны быть общедоступными (open license). В докладе приводятся результаты исследования проведенного Немецкой национальной библиотекой, около 70 % баз данных которой уже доступны в виде открытых связанных данных.

¹ В докладе под этим термином понимается возможность осуществления поиска публикаций в национальной библиографии, т.е. использование национальной библиографии в качестве справочного издания

Введение

Национальную библиографию можно определить как исчерпывающий список публикаций ограниченной географической области². В эпоху Интернета нельзя придерживаться этого определения, т.к. Интернет изменил оба понятия: и «публикация» и способ опубликования.

Традиционная печатная книга самодостаточна и постоянна, а Интернет-публикации не имеют четко очерченных границ, они динамичны и интерактивны, их содержание постоянно меняется. Кроме того, изменилась техника записи текстовой продукции: с появлением полнотекстового поиска каталоги, базирующиеся на метаданных, тезаурусы и классификационные системы превратились в изолированное сообщество. Это не означает, что полнотекстовый поиск качественнее, чем поиск на основе метаданных, но его проще автоматизировать: полнотекстовые ресурсы легко доступны в Интернете, в то время как высококачественные метаданные необходимо еще создать. Создание метаданных – или каталогизация – возможно либо на основе комплексной автоматизации процессов либо на основе человеческого труда.

Но сейчас многие процессы, основанные на человеческом труде, происходят за пределами библиотек: издатели для усиления своих позиций используют собственные метаданные и такие онлайн платформы, как Википедия или OpenLibrary, предлагают общественности публиковать статьи и метаданные, описывающие эти публикации, создавая тем самым идеальное дополнение к полнотекстовому поиску. Недавний рост массовой оцифровки и оптического распознавания символов изменили ситуацию, предоставив еще более широкую возможность осуществлять полнотекстовой поиск в (ранее) неоцифрованных материалах. Для поисковых систем, таких как Google, это оказалось более эффективным, чем использовать традиционные методы каталогизации; в поисковой системе поиск на основе метаданных является просто дополнительной функцией³.

Учитывая это, библиотеки должны спросить себя, какую дополнительную услугу сейчас и в будущем может предоставить национальная библиография, базирующаяся на традиционной каталогизации.

² Anderson (1974), p.12

³ <http://books.google.com/intl/en/googlebooks/about.html>

Требования к национальной библиографии

Традиционно пользователи национальной библиографии делятся на три группы: издатели и книготорговцы, библиотеки и конечные пользователи (в частности, исследователи и литературоведы). Общим для этих трех групп является то, что для них важны следующие четыре основных свойства национальной библиографии:

1) Полнота и достоверность данных

Для книготорговцев, издателей и исследователей всегда было важно, чтобы в национальной библиографии максимально полно отражалась вся продукция без каких-либо политических или содержательных ограничений. Кроме того, не маловажную роль играло соответствие правилам каталогизации, особенно при создании кумулятивной библиографии.

2) Оперативность данных

Для книготорговцев и библиотек была особенно важна оперативность представления данных в национальной библиографии. Во второй половине 20-го столетия с целью сокращения цикла издательской подготовки много внимания уделялось обработке данных, и это несмотря на рост количества публикаций.

3) Репрезентативность (Referentiability)

Обладая полнотой, достоверностью и оперативностью национальная библиография могла выполнять справочные функции при проведении научных исследований: если книга была учтена в национальной библиографии, то можно было быть уверенным, что она существует, и если книги нет в национальной библиографии, скорее всего, она никогда не была издана.

4) Стабильность

Для выполнения справочных функций было недостаточно, чтобы записи соответствовали приведенным выше критериям качества, нужна была еще и стабильность. До начала 21-го века проблемы не существовало, так как библиография представляла собой печатное издание. Можно было только исправить неправильные библиографические записи в кумуляции, но было невозможно удалить сведения об опубликовании книги.

Требования, предъявляемые к национальной библиографии, возрастают и распространяются теперь на используемые ею данные и повторно используемые во всемирной паутине.

Интернет все больше превращается в открытое пространство по обмену данными: так называемое облако связанных данных⁴. С 2008 года эта коллекция взаимосвязанных наборов данных значительно увеличилась, но мало что известно о качестве и стабильности данных. Для функционирования семантической сети необходим определенный уровень надежности, а также качества и, особенно, стабильности информации. Эта сеть может устойчиво развиваться только при условии, что информация, которую мы проаннотировали сегодня, будет доступна и на следующей неделе: чтобы было на что ссылаться (citeable)⁵. Эта проблема, в частности, касается онлайн публикаций и их метаданных. Для обеспечения долгосрочной надежности необходимо стабилизировать эту часть Интернет. В будущем это должно стать задачей библиотек и других организаций, сохраняющих культурное наследие страны.

Вывод:

1) Полнота и достоверность данных

Рост онлайн-публикаций делает практически невозможным отражение в национальной библиографии всей созданной продукции. Скорее всего, национальное библиографическое агентство не сможет даже точно ответить на вопрос о полноте отражения публикаций в национальной библиографии. Вместо этого, оно должно будет определить, какие типы публикаций будут отражаться с наибольшей полнотой (например, печатные средства массовой информации, высококачественные блоги, некоторые типы научных монографий), и где будут храниться периодически сканируемые совокупности веб-сайтов.

Некоторые метаданные будут создаваться в автоматизированном режиме, и они будут далеки от совершенства. Но это не значит, что их нельзя будет использовать, процессы должны быть хорошо задокументированы, и потребители должны четко видеть, что эти данные были созданы машиной и их следует использовать с учетом способа создания.

Мы должны придерживаться определенных требований и меньше уделять внимания таким правилам каталогизации как AACR⁶ или RAK⁷, больше обращать внимание на технические и семантические данные взаимодействия для дальнейшей интеграции во внешние сервисы.

⁴ Heath and Bizer (2011)

⁵ Schuster and Rappold (2006)

⁶ Anglo-American Cataloguing Rules, <http://www.aacr2.org/>

⁷ Regeln für die alphabetische Katalogisierung,

http://www.dnb.de/EN/Standardisierung/Regelwerke/regelwerke_node.html#doc3132bodyText1

2) Оперативность данных

Многие клиенты ожидают, что данные об интернет-изданиях появятся в национальной библиографии в момент их опубликования. Но это противоречит требованиям репрезентативности и стабильности: если национальная библиография базируется на метаданных, описанных ранее версий публикации, то существует высокий риск внести изменения в библиографическое описание. Для этого библиотечно-информационные системы должны иметь возможность архивировать метаданные различных версий, сохраняя их разный статус.

3) Репрезентативность

Национальная библиография и в будущем останется основным источником выполнения справок о публикациях. Особенно это касается электронных публикаций, т.к. их роль будет только возрастать. Для обеспечения этой услуги, необходимы уникальные идентификаторы библиографических данных, функционирующие в электронной среде.

4) Стабильность

Долговременная репрезентативность невозможна без долговременной доступности и гарантированной полноты библиографических описаний. Это имеет значение не только для надежной идентификации библиографических и авторитетных данных, но это и один из способов осуществления контроля, внесения исправлений и удаления в ручном режиме, выявления определенных версий, например, при помощи меток.

Топология немецкой национальной библиографии в цифровую эпоху

Учитывая эти требования, национальная библиография не должна отделять себя от всемирной паутины. В Германии национальная библиография превратилась из набора каталожных карточек в (электронную) базу данных, использующую все более и более сложные структуры данных, в которых библиографические описания связаны с авторитетными данными, а часто и с ресурсами вне системы. Следующим шагом станет вхождение национальной библиографии во всемирную паутину, и каждый ее аспект станет неотъемлемой частью Интернета: в топологическом, функциональном, техническом и организационном плане.

Проанализировав эти характеристики, мы осознали нашу национальную библиографию как граф во всемирной паутине. Этот граф имеет внутренние связи, а также связи с другими частями Интернета. Представление библиотечных данных в виде графа в сети

возможно не только по отношению к немецкой национальной библиографии. Это могут быть указатели других объектов культурного наследия, авторитетные файлы, тезаурусы и классификационные системы.

В будущем национальная библиография не сможет оставаться закрытой системой, она многое потеряет, если будет игнорировать веб. Полнота национальной библиографии невозможна без учета данных, находящихся вне библиотечных баз данных.

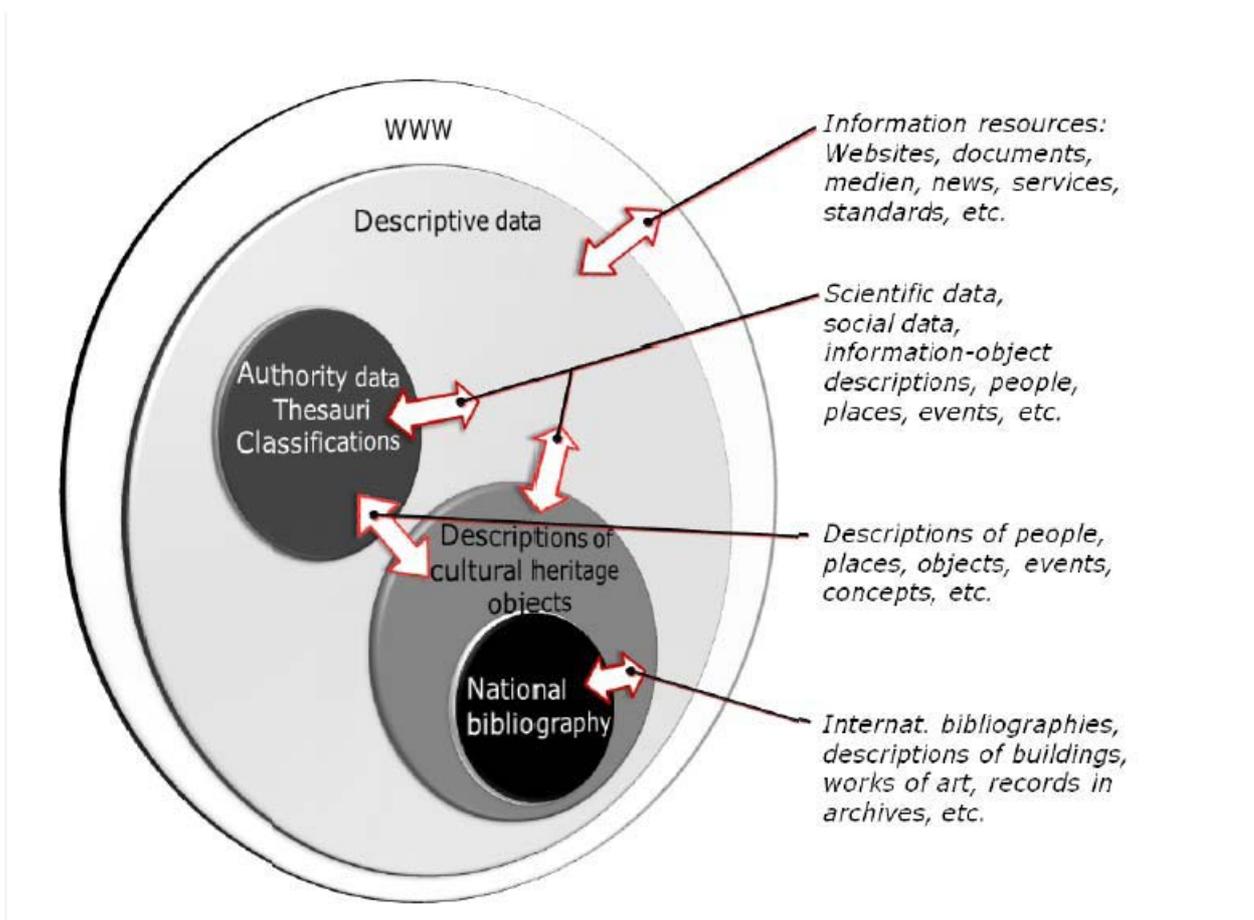


Рисунок 1: В будущем национальная библиография может быть смоделирована в виде графа во всемирной паутине. Эллипсы – подграфы. Стрелки между подграфами показывают возможные взаимосвязи между узлами внутри подграфов.

Границы и ограничения

Трудно определить, где именно национальная библиография начинается и где заканчивается. Существуют границы, имеющие отношение к географии региона (страны) и определенным типам объектов культурного наследия (например, книги, карты,

открытки, звукозаписи и т.д.)⁸. Также существуют границы административной ответственности национальной библиографии.

Для удовлетворения требований долговременной надежности и высокого качества данных орган, отвечающий за национальную библиографию, должен иметь соответствующую государственную поддержку и соответствующее оборудование для управления и обработки данных. В этом заключается основной смысл деятельности национальной библиотеки или национального библиографического агентства, опирающейся – но не всегда – на законодательство об обязательном экземпляре⁹. Из этого не следует, что все данные, хранящиеся в национальном библиографическом агентстве, автоматически становятся частью национальной библиографии: Немецкая национальная библиотека поддерживает несколько баз данных, в том числе специальных коллекций, которые не имеют никакого отношения к немецкой национальной библиографии.

Концепции

Во время семинара в Немецкой национальной библиотеке участники должны были определить суть немецкой национальной библиографии. Ответы варьировались от "все данные, хранящиеся в библиотеке" до "только заглавия и исключая специализированные фонды". Тем не менее, даже группа, выступающая за минимум, не была готова полностью отказаться от авторитетных данных. Это не удивительно, т.к. авторитетные файлы являются продуктом процесса каталогизации и содержат информацию, например, имя автора, и без которой невозможно создать библиографическое описание в соответствии с ISBD, то есть без этой информации библиографические записи не будут полными. С другой стороны, в печатной национальной библиографии при создании библиографических описаний на основе ISBD используются не все данные из авторитетной записи.

Получается, что невозможно ответить на вопрос, какие данные являются основой национальной библиографии не только на уровне объекта, но и на уровне индивидуальных характеристик¹⁰. В идеале для оптимизации процесса поиска онлайн каталог должен задействовать все имеющиеся данные. Это могут быть данные из внебиблиотечной сферы, созданные без каких-либо ссылок на библиографические объекты. Одним из примеров является использование географических координат из такой

⁸ IFLA (2009)

⁹ Andersen (1974), p. 11

¹⁰ IFLA (2009), p.4

базы данных, как geonames.org, что позволяет нам осуществлять поиск по географическому местоположению или географической близости.

Вывод библиографического описания, соответствующего ISBD, и поиск в онлайн-каталоге являются лишь частью услуг, предоставляемых национальной библиографией. Не все данные задействованы при оказании этих двух услуг и, следовательно, мы не можем, принимая во внимание только эти услуги, определить, что составляет самую суть национальной библиографии (печатные библиографии, поиск в Интернете, национальные библиографические услуги).

Когда (национальные) библиографические службы работают не только с данными, используемыми в национальной библиографии, то важно не преувеличивать роль этих данных. Четкая и последовательная картина будет тогда, когда мы сосредоточим свое внимание на сути национальной библиографии: публикациях. В этом случае мы будем иметь минимальный набор данных, состоящий из текстовых элементов и ссылок, которые возможно сами по себе не представляют большой ценности. Ориентируясь на FRBR¹¹ и RDA, можно сказать еще более точно: национально-библиографические записи воплощений это "физическое воплощение выражения произведения"¹². Данные о произведениях, выражениях и физических единицах – как авторитетные записи – являются элементами их собственных наборов данных, а данные в национальной библиографии содержит только ссылки на эти объекты.

Процессы создания

Национальные библиографии должны стать частью глобальной экосистемы данных потребителей и данных производителей. Чем сильнее будет их связь с доменными независимыми данными, создаваемыми широким кругом учреждений, тем будет выгоднее для них, т.к. они будут получать сведения из этих баз знаний. В будущем библиотечное обслуживание все больше будет опираться на данные, полученные из внешних источников. Наличие дополнительной информации об объектах, имеющих отношение к культурному наследию – культурные ценности, субъекты или концепции, позволяет улучшить наше библиографическое обслуживание. Возможно, мы откажемся от хранения у себя некоторой части данных, а взамен будем повторно использовать информацию, созданную третьими лицами.

¹¹ FRBR <http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>

¹² IFLA (1998), p. 21.

Это не новое решение вопроса. Уже сегодня в Немецкой национальной библиотеке мы сами вручную не составляем все библиографические описания, а повторно используем информацию, созданную третьими лицами, например, метаданные, поступающие от издателей, или информация, созданная в полуавтоматическом режиме – оглавления; или созданные машинным способом предметные рубрики. Скорее всего, следующим шагом станет увеличение количества повторно используемых данных и открытая контролируемая каталогизация избранных сетевых ресурсов.

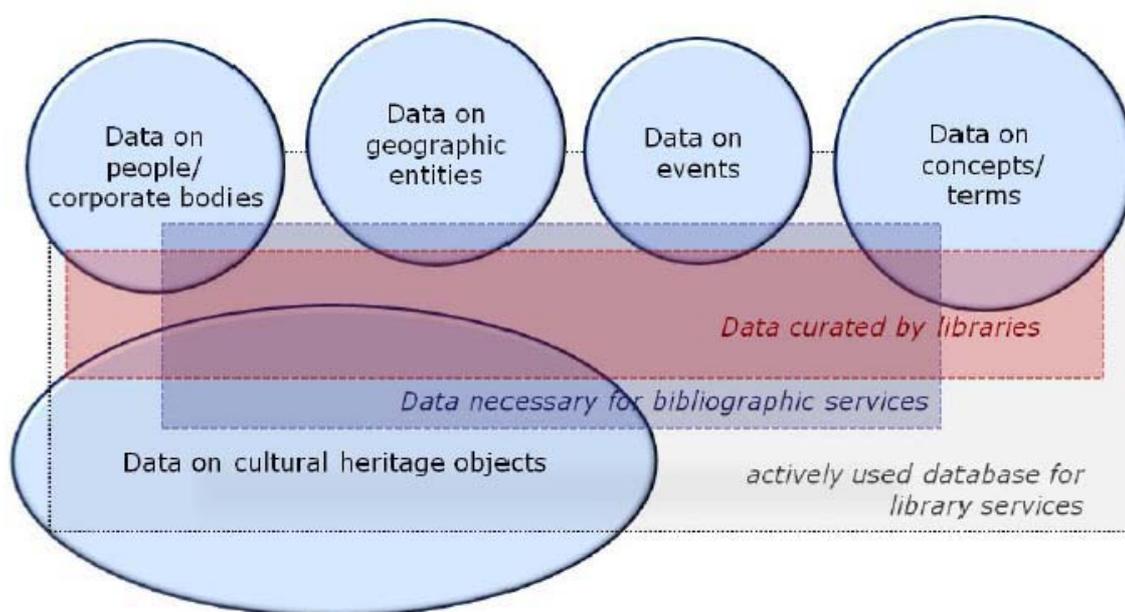


Рисунок 2: Проект распространения созданных и используемых в библиотеках данных.

1. Не все компоненты библиотечных каталогов используются при библиографическом обслуживании
2. Не все компоненты данных, используемые при библиографическом обслуживании, созданы/закаталогизированы библиотеками.
3. Количество используемых библиотеками баз данных больше, чем набор данных, состоящий из двух частей.

Это позволит библиотекам отказаться от выполнения некоторых задач и сконцентрировать внимание на том, что они делают лучше других: создание высококачественных данных, организация знаний и осуществление авторитетного контроля. В результате каждый будет работать более эффективно и будет иметь возможность сконцентрироваться на выполнении своей задачи, и все только выиграют от

этого. Национальный библиографический центр (если нет другой организации, создающей подобную же качественную продукцию) отвечает за опубликованные объекты культурного наследия в текстовой форме, за все ссылки от этих объектов к другим объектам и за их свойства, без которых невозможно библиографическое обслуживание. Другие поставщики данных отвечают за другие свойства.

В связи с увеличением числа публикаций, библиотекам придется сократить создание, поддержку и проверку метаданных в ручном режиме, т.к. создаваемая в результате этих процессов добавленная стоимость не выдерживает конкуренции с данными, созданными в автоматическом режиме. В этом случае определение добавленной стоимости не будет зависеть от соблюдения определенных правил каталогизации, вместо этого надо, с одной стороны, сосредоточиться на повышении качества отбора данных и фильтрации, а с другой стороны – на бизнес-моделях библиотечного обслуживания клиентов.

Характеристики качества

Качество библиотечного обслуживания не должно страдать от привлечения услуг третьих лиц. Наоборот – из этого можно извлечь выгоду, только надо тщательно изучить внешние источники и определить какую информацию можно использовать повторно. Надежность (репрезентативность, постоянство, использование стандартов и гарантия качества) данных имеет решающее значение. Происхождение данных и – по-возможности – процесс их создания, должны быть прозрачны, и проходить качественный контроль на месте. Каждый внешний набор данных должен проходить экспертизу на возможность его долгосрочного использования.

Повторно используемые в национальной библиографии наборы данных от третьих лиц могут не соответствовать существующим правилам каталогизации. Много – и даже очень много – поставщиков данных, не входящих в библиотечное сообщество, не имеют представления об AACR II, RAK или RDA, и мы не можем ожидать, что они будут осваивать их. Вместо этого мы должны упростить существующие правила каталогизации и оценить их с точки зрения такого вопроса «Как они помогают национальной библиографии выполнять ее задачи: описать с исчерпывающей полнотой все вышедшие в стране публикации». Если правила с этим справляются, то мы должны их использовать при оценке качества наборов данных, поставляемых третьими лицами.

Если организация использует метаданные, поставляемые третьей стороной, то она должна убедиться в надежности поставщика данных. Для этого надо составить перечень критериев, например, аналогичный перечню, применяемому в надежных цифровых

архивах. Главным критерием в списке будет критерий совместимости данных. Данные, не поступающие от надежных поставщиков, должны быть особо отмечены.

То же самое относится и к данным, создаваемым в автоматическом режиме. Здесь важно, чтобы процессы были хорошо задокументированы, а результаты воспроизводимы. Если мы сможем создавать однородные данные для подмножеств национальной библиографии, то доверие клиентов будет только расти. Это станет возможным, если совместно будет составлен перечень критериев надежности для автоматизированных процессов. Если такой список будет, то контроль над процессами будет осуществляться на регулярной основе.

Техническое воплощение

Выше говорилось, что национальная библиография может базироваться на сочетании минимального набора метаданных воплощения, информации из авторитетных файлов и взаимосвязанных наборов данных. Из этого, а также из проекта представления библиографии в Интернете в виде графа, вытекает естественный вывод использовать методику связанных данных – по сути, основанную на протоколе http для машиночитаемых связей между объектами – для технического воплощения библиографии.

Учитывая требования Tim Berners-Lee для связанных данных¹³, реализовать это будет просто:

- 1) Использовать URI (универсальный идентификатор ресурса) в качестве имен объектов

Каждый объект в национальной библиографии должен иметь свой собственный уникальный идентификатор. Это относится, по крайней мере, к воплощениям, словарям контролируемых терминов, коллективным и индивидуальным авторам; а в идеале и к другим авторитетным данным, таким как издатели.

- 2) Использовать HTTP URI таким образом, чтобы эти имена можно было найти

При присвоении идентификаторов своим объектам, библиотеки должны использовать инфраструктуру Интернета.

- 3) Конечным результатом поиска URI должно стать получение ценной информации

Для получения максимальной пользы от повторного использования библиотечных данных, важно предоставить краткие описания данных узла, найденных по URI.

Делая так, мы приглашаем других поставщиков данных, используя наши URI,

¹³ Berners-Lee (2006)

повторно использовать совокупности библиотечных данных вместе с авторитетными данными или библиографической информацией.

- 4) Для получения большего количества информации включать ссылки на другие URI
Когда мы для наших библиографических записей повторно используем авторитетные данные или данные других контент-провайдеров, мы имеем возможность выбирать *некоторые* свойства, связанные с объектом, например, имя автора или предпочтительная форма предметной рубрики, заглавие воплощения, а также указать ссылку на страницу, на которой пользователь может найти дополнительную информацию: дата рождения или профессия.

При создании такой инфраструктуры не важно, смоделированы данные на основе RDF (Resource Description Framework) или других технических средств, связанные данные сами могут быть в формате агностик. RDF стек-технология обладает большими возможностями создания в общем виде машиночитаемых (а, следовательно, и машинопонимаемых) данных и семантических связей. Часто это приводит к очень сложной модели данных и проблемам, которые могут возникнуть при использовании RDF для развернутых приложений. Такая практика может способствовать возникновению острой дискуссии¹⁴.
Немецкая национальная библиотека имеет подобный опыт. Пока мы последовательно строим на принципах связанных данных наши модели данных – выставляя наши данные в Интернете под стабильными идентификаторами и снабжая их ссылками на другие объекты – мы прокладываем путь для введения более мощных языков описания.

Использование

При поиске информации в Интернете, большинство пользователей обращаются к коммерческим поисковым системам, таким как bing, ask.com и Google или социальным сетям LinkedIn и Facebook¹⁵. Для (национальных) библиотек и других учреждений, имеющих дело с культурным наследием, это означает, что они должны предоставить этим компаниям возможность творчески интегрировать наши данные в свои услуги, а в результате пользователи найдут в Интернете необходимую и полезную для них информацию. Например, портал европейского культурного наследия Europeana пришел к выводу, что создание специальной собственной страницы для каждого объекта коллекции

¹⁴ Miličić (2011)

¹⁵ Согласно докладу PEW, 92 % взрослых пользователей сети ищут информацию при помощи поисковых машин. Purcell (2011).

и присвоение этой странице уникального URI, улучшит видимость этого объекта в сети и увеличит трафик из поисковых систем¹⁶.

Это означает, что библиотечный онлайн-каталог будет продолжать играть важную роль в информационном пространстве. Узкоспециализированные поисковые системы являются эталоном, показывающим, что можно сделать с библиотечными данными и как эти данные могут быть адаптированы к запросам конкретных групп пользователей. С другой стороны, граница между добавленной стоимостью, предлагаемой библиотечным каталогом и специализированной поисковой системой Интернет, становится все более размытой. Мы можем представить, что в будущем некоторые провайдеры поисковых систем сфокусируют основное внимание на библиографической информации, например, для конкретных научных сообществ. Наличие высококачественных библиографических данных, встроенных в среду, в которой осуществляется поиск информации, будет полезно для наших постоянных клиентов и облегчит поиск, отбор и получение необходимых им ресурсов.

Для того чтобы другие службы могли легко повторно использовать библиотечные данные, необходимо четко указывать условия лицензирования. Именно в национальной библиографии добавленная стоимость создается посредством увеличения прозрачности и повторного использования данных, и достичь это можно очень простым способом: предлагать данные на основе гибких и открытых лицензий, разрешая коммерческое повторное использование. Это может быть спорным решением, учитывая, что такие крупные игроки как Google в состоянии платить за библиотечные метаданные. Однако, сомнительно, что они будут платить, и если библиотечная информация не будет отображаться в основных поисковых системах, то она будет невидима для большинства людей, осуществляющих поиск в Интернете. Использование открытой лицензии также позволяет интегрироваться в такие открытые платформы, как Википедия, в малые и средние веб-сайты, например, новостные Интернет-порталы. Таким образом, повышение видимости в Интернете обязательно перевесит (потенциальную) потерю дохода.

Текущая деятельность Немецкой национальной библиотеки и ее последствия для немецкой национальной библиографии

Немецкая национальная библиотека, как и другие национальные библиотеки, постоянно совершенствует свою деятельность.

¹⁶ Clark et. al. (2011), p. 15-17.

Начиная с 2008 года Немецкая национальная библиотека последовательно выставляет свою базу данных в виде открытых связанных данных; и сейчас предоставлен доступ приблизительно к 70 % данных о заглавиях. Первыми были авторитетные данные GND, затем немецкий перевод десятичной классификации Дьюи. Спустя шесть месяцев с начала запуска проекта, мы смогли предоставить в виде связанных данных основную часть данных заглавий национальной библиографии. Кроме данных DDC, которые доступны по лицензии CC BY-NC-ND, все остальные данные Немецкой национальной библиотеки опубликованы на базе Creative Commons Zero.

Кроме того, мы все чаще создаем метаданные описаний и предметов в автоматизированном режиме. В первую очередь, это позволяет дополнить каталожные библиографические описания, например, оглавлениями, и мы стали больше использовать метаданных от издателей и университетов, в том числе и предметные категории, например, BISAC коды и предметные рубрики. Для онлайн публикаций в серии О национальной библиографии теперь применяются только повторно используемые метаданные описаний и предметов, поставляемые самими владельцами контента¹⁷.

Сейчас мы оцениваем возможность автоматического добавления предметных рубрик в библиографические записи. Этот процесс построен на основе автоматического индексирования. Мы анализируем возможность обработки различных типов документов и качество поиска публикаций на основе автоматического индексирования. Первые результаты будут доступны в 2013 году.

Автоматическое индексирование веб-ресурсов будет осуществляться на основе тех же контролируемых авторитетных рубрик, какие мы используем при интеллектуальном индексировании, чтобы было можно осуществлять поиск однородной информации. Таким образом, мы предлагаем надежные авторитетные данные. Кроме того, можно будет эффективно использовать связанную информацию о людях, местах и т.д., а также семантические связи в авторитетных данных.

Пользователи обращаются к библиографическим данным, преследуя разные цели, в первую очередь их интересует предметный доступ, предоставляющий обзор имеющейся литературы, библиографические ссылки или прямой доступ к публикациям. Немецкая национальная библиотека проводит последовательную политику индексирования и использования контролируемого доступа, т.к. они остаются важными инструментами, обеспечивающими порядок и согласованность данных. Такая практика соответствует рекомендациям ИФЛА¹⁸. Отраслевое (предметное) расположение материала в

¹⁷ Gömpel and Svensson (2011)

¹⁸ IFLA (2012)

национальной библиографии наилучшим образом соответствует запросам пользователей, такое расположение также позволяет не обращать внимания на формат данных, каналы распространения или вывод информации на экран. Классификационные системы и предметные рубрики помогают пользователям получить (найти, идентифицировать, отобразить или проанализировать) ту информацию, которую они хотят получить. Большинству пользователей (конечным пользователям или профессионалам) нужна не полная библиография, а только небольшие списки. Многие пользователи могут даже не догадываться, что, обращаясь к нашему онлайн каталогу, они ищут в библиографии. Поиск в библиографии будет еще удобнее, если настроить личные профили или подписаться на RSS рассылку результатов поиска. Наличие дополнительной информации, например, оглавлений или рефератов, помогают пользователям лучше ориентироваться в информации. Предоставление связанного онлайн-контента, особенно из электронных книг и электронных журналов, отвечает потребностям пользователей.

В связи с растущим числом опубликованной информации становится все более важным классифицировать информацию по управляемым модулям, так чтобы можно было ее прочесть, отобразить и осуществить точный поиск. В идеале, полное и подробное индексирование всех документов позволяет осуществлять релевантный тематический поиск или поиск дополнительной содержательной информации. Огромное количество печатных публикаций и все возрастающее число опубликованных веб-ресурсов в немецком домене, а также сокращение бюджета и персонала, заставили нас пересмотреть политику индексирования: мы используемый разный уровень индексирования национальной продукции. Эта модульная политика¹⁹ предусматривает использование разных точек доступа, отбор разных типов СМИ, но при этом все по-прежнему остается прозрачным и находится под контролем критериев качества.

Минимальный уровень индексирования – это очень широкий классификационный индекс, который присваивается почти всем ресурсам – *Sachgruppe* / предметная группа, эквивалентная общему классификационному индексу. Структура построена на основе двух верхних иерархических уровней DDC (сто отделов). Для удовлетворения некоторых запросов применяется более детальное индексирование. Фантастика, книги для детей, а также школьные учебники отбираются выборочно.

Некоторые другие проекты направлены на улучшение связи с другими наборами данных. Отличной отправной точкой является оптимизация наших собственных данных. Автоматизация позволяет отказаться от текстовых описаний и начать использовать ссылки. Ссылки на внешние наборы данных Немецкая национальная

¹⁹ Svensson and Jahns (2010)

библиотека создает и в ручном, и автоматизированном режимах. Через VIAF (Международный виртуальный авторитетный файл) мы смогли автоматически связать лицо из GND с другими авторитетными файлами, а через MACS²⁰ с предметными рубриками из GND, LCSH и RAMEAU. Сотрудничество с немецкой Википедией обеспечивает связь между статьями Википедии и лицами в GND, а в CONTENTUS мы разработали методику создания сети географических названий из GND, статей из Википедии и geonames.org.

В целях дальнейшего развития Немецкая национальная библиография совместно с HBZ, библиотечной сетью Северный Рейн – Вестфалия создала платформу culturegraph.org. Основная идея заключается в создании технической и организационной инфраструктуры для поддержания взаимосвязанных библиографических данных, тезаурусов, классификационных систем и других авторитетных данных. В результате информационная сеть будет представлена в виде связанных данных. Одним из основных атрибутов платформы является поддержание возможности ссылаться (citeability). В настоящее время платформа содержит библиографические описания всех публикаций, начиная с 1945 года, хранящихся в немецкой библиотечной сети. Промежуточной целью является создание кластеров данных позволяющих определить, какие библиографические записи описывают одно и то же воплощение, присвоить имя и опубликовать эти кластеры, используя общий уникальный идентификатор. Дальнейшее развитие покажет, можно ли создать кластеры для произведений (в соответствии с FRBR).

Национальная деятельность должна строиться с учетом международных тенденций развития, и, возможно, наиболее важной задачей станет дальнейшее участие в международных проектах и рабочих группах по разработке стандартов по каталогизации и обмену данными.

Выводы и будущая деятельность

В докладе мы обосновываем создание экосистемы библиотечных данных на основе принципов связанных данных, выставление их в Интернет на условиях открытой лицензии. Для более широкого распространения этих данных, мы должны сотрудничать с провайдерами крупных поисковых систем и иметь возможность интегрировать эти библиографические данные в их базы данных.

²⁰ Многоязычный предметный вход.

http://www.nb.admin.ch/nb_professionnel/projektarbeit/00729/00733/index.html?lang=en

Для максимально полного удовлетворения самых разнообразных потребностей клиентов – отбор данных, форматы, – необходимо чтобы основная база данных обладала гибкими возможностями. Что касается индексирования, то мы должны перейти от предкоординатных данных к информационной архитектуре, в которой мы можем посткоординировать биты и фрагменты информации с учетом постоянно меняющихся требований.

В правилах каталогизации следует больше внимания уделять сетевым библиографическим данным, т.к. они становятся все более востребованными. RDA стремятся стать "новым стандартом описания ресурсов и доступа, предназначенным для цифрового мира"²¹. RDA справедливо критикуют за то, что они по-прежнему ориентированы на традиционные критерии каталогизации и не сделали значительного рывка вперед, но все же это шаг в правильном направлении, если учитывать их удивительную гибкость. Сомнений в необходимости проведения дальнейших изменений быть не может.

Может быть, более важным, чем библиотечные правила каталогизации, является возможность взаимодействовать с данными из внебиблиотечных экосистем. Много доступных и полезных данных – например, издательская реклама новых изданий, содержащая слова, на основе которых можно вести поиск – не учитываются при поиске и отборе. Надо, чтобы библиотеки имели возможность хранить данные, созданные третьими лицами, и делать эту информацию прозрачной для конечных пользователей.

Далее необходимо сделать возможность ссылаться (citeability) критерием качества. Если библиографическое описание будет снабжено пометкой "citeable", то его никогда нельзя будет удалить, и в дальнейшем можно будет проследить все изменения, происходившие с ним. Для этого надо активизировать процессы создания и хранения данных, включить во все данные, созданные в автоматическом режиме, сведения о происхождении информации.

Наконец, мы должны вкладывать больше средств в развитие возможности включения библиотечных данных в другие наборы данных. По-возможности, мы должны использовать ссылки на авторитетные данные, а не создавать текстовый контент ("literals"). Международное сотрудничество по общим авторитетным файлам – например, VIAF – является важным шагом, позволяющим избежать двойной работы. Центральное хранилище заглавий произведений окажет неоценимую помощь в использовании библиографических описаний, не взирая на национальные границы и языки. База данных издателей (подмножество коллективных авторов) не только пополнит контролируемые точки доступа, но также будет способствовать сотрудничеству библиотек и издателей.

²¹ The strategic plan for RDA 2005-2009. <http://rda-jsc.org/stratplan.html>

Национальная библиография должна стать одним из графов среди сетевых графов. Цифровая эпоха изменила информационный сектор. Вместо того, чтобы отстаивать существующие структуру и практику, библиотеки должны принять вызов и выступать в роли агентов перемен. Для этого нужны большая гибкость и открытость.

Ссылки

Anderson (1974) Anderson, Dorothy: Universal bibliographic control : a long term policy; a plan for action. Pullach 1974.

Berners-Lee (2006) Berners-Lee, Tim: Linked Data. Available at <http://www.w3.org/DesignIssues/LinkedData.html>

Clark et. al (2011) Clark, D. J.; Nicholas, D.; Rowlands, I.: D3.1.3 – Publishable report on best practice and how users are using the Europeana service. Available at http://www.europeanconnect.eu/documents/D3.1.3_eConnect_LogAnalysisReport_v1.0.pdf

Gömpel and Svensson (2011) Gömpel, Renate; Svensson, Lars G.: Managing Legal Deposit for Online Publications in Germany. 2011. urn:nbn:de:101-2011061609. Online available at <http://nbn-resolving.de/urn:nbn:de:101-2011061609>

Heath and Bizer (2011) Heath, Tom; Bizer, Christian: Linked Data: Evolving the Web into a Global Data Space. San Rafael (Calif.) 2011. (Also online at <http://linkeddatabook.com/editions/1.0/>)

IFLA (1998) Functional requirements for bibliographic records : final report. München 1998. All references in this paper are to the 2009 revised online version at http://www.ifla.org/files/cataloguing/frbr/frbr_2008.pdf

IFLA (2009) IFLA Cataloguing Principles: Statement of International Cataloguing Principles (ICP) and its Glossary. Ed. Barbara Tillett and Ana Lupe Christián. München 2009. For a list of online versions (incl. translations) see <http://www.ifla.org/publications/ifla-series-on-bibliographic-control-37>

IFLA (2012) , Guidelines for Subject Access in National Bibliographies. Ed. Yvonne Jahns. Berlin: 2012.

Miličić (2011) Miličić, Vuk: The Ultimate Problem of RDF and the Semantic Web. Blog post, available at <http://milicicvuk.com/blog/2011/07/19/ultimate-problem-of-rdf-and-semantic-web/>

Purcell (2011) Purcell, Kristen: Search and email still top the list of most popular online activities: Two activities nearly universal among adult internet users. 2011. Available online at http://pewinternet.org/~media/Files/Reports/2011/PIP_Search-and-Email.pdf

Schuster and Rappold (2006) Schuster, Michael; Rappold, Dieter: Social Semantic Software – was soziale Dynamik im Semantic Web auslöst. In: Semantic Web: Wege zur vernetzten Wissensgesellschaft. Ed. Tassilo Peregrini. Berlin 2006.

Svensson and Jahns (2010) Svensson, Lars G.; Jahns, Yvonne: **PDF, CSV, RSS and other acronyms: redefining the bibliographic services in the German National Library**. 2010. Available online at <http://www.ifla.org/files/hq/papers/ifla76/91-svensson-en.pdf>

Перевод с сокращениями А.В. Теплицкой
(Российская государственная библиотека, Москва)