



Заглавие: Агрегация данных и распространение авторитетных данных через Linked Open Data¹

Авторы²:

Xavier Ajenjo, Менеджер проекта Fundación Ignacio Larramendi и директор виртуальной библиотеки «Полимат»
[xavier.agenjo@larramendi.es]

Francisca Hernández, Директор по метаданным. Консультант. DIGIBÍS, Producciones Digitales
[francisca.hernandez@digibis.com]

Viedma, Andrés, Архитектор ПО. DIGIBÍS, Producciones Digitales
[andres.viedma@digibis.com]

Translated by: Aya KHLEBAEVA
National Library of Russia
Saint Petersburg

Meeting:

**80 - Каталогизация: разрушение барьеров - Секция
Каталогизации**

Резюме:

Агрегация данных и распространение авторитетных данных через Linked Open Data описаны путём анализа Виртуальной библиотеки полиматов (Polymath Virtual Library). Её задача – собрать в одном месте информацию, данные, оцифрованные тексты и веб-сайты, посвящённые испанским, испано-американским, бразильским и португальским полиматам всех времен. Поэтому она агрегирует информацию о мышлении, философии, политике, науке и пр. в Испании, испаноговорящей Америке, Португалии и Бразилии, написанную на любом языке (латынь, арабский, древнееврейский, испанский, португальский) и в любое время (начиная с Сенеки в первом веке до н.э. и до настоящего времени). Стержнем этой системы стали авторы. На каждого автора создана авторитетная запись (A3) MARC21/RDA, которая обогащена биографическими данными. Определены специфические атрибуты для расширения связей и возможности навигации по сайту (профессия, род занятий, пол, членство в различных организациях, даты рождения и смерти, места рождения и смерти, а также – использование языков и графики). Таким образом каждая A3 агрегирует информацию из многочисленных разнообразных источников и словарей. Кроме того, электронные версии произведений каждого автора описаны по MARC/RDA (12

¹ Все цифры и URL сверены и обновлены по состоянию на 30 апреля 2011 г.

² Авторы благодарят за редактирование доклада César Juanes, DIGIBIS.

версия). Схожим образом каждый автор имеет связь с другими авторами (переводчиками, издателями, комментаторами и т.д.), - в качестве текстуальной передачи их произведений с использованием АЗ «автор-заглавие», которые, согласно последнему решению MARBI, построены по модели «Произведение – выражение – воплощение – физическая единица» (или **WEMI**). Эти АЗ созданы после наведения справок в различных проверенных источниках (авторитетных файлах, энциклопедиях, биографических словарях и т.п.). До 2010 г. связь между конкретным автором и найденными данными устанавливалась вручную. С 2011 года Фонд Игнасио Ларраменди (*Fundación Ignacio Larramendi*) разрабатывает различные инструменты для того, чтобы в полуавтоматизированном режиме получать данные из источников *Linked Open Data*. Доклад описывает детали получения URI ресурсов LOD, генерирующих автоматические запросы на серверах *SRU/OpenSearch* и *SPARQL Endpoint*, или выполняющих это семантическое обогащение через файлы, доступные на LOD. Экспорт данных производится в *MARC 21*, *Europeana Data Model 5.2.1*, *SKOS* и *VIAF*, после семантического обогащения в таких словарях как *Список предметных рубрик Библиотеки Конгресса (ПР БК)*, *Виртуальном авторитетном файле, GeoLinkedData (Национальный географический институт)* и *GeoNames*.

1. Введение: Виртуальная библиотека «Полимат»

Следующие далее сведения об агрегации и распространении АЗ в *Linked Open Data* основаны не на теории, а на современной библиотечной практике, - так каталогизируют в Виртуальной библиотеке «Полимат», являющейся составной частью Виртуальных библиотек Фонда Игнасио Ларраменди³. основополагающая задача Виртуальной библиотеки «Полимат» состоит в том, чтобы показать особую значимость и важность испанских, португальских, бразильских и латиноамериканских мыслителей (полимагов) и, - что является одной из основных отличительных черт этого библиографического проекта, - связать их с другими произведениями, имеющими схожие характеристики, открыв пользователям всего мира доступ к этому богатству.

Обзор начинается с обозначения значения прежних библиографических разработок, которым следует отдавать предпочтение перед вопросами технологии, особенно – в таком новом начинании, каким является *Linked Open Data*. Именно эти библиографические задачи являются основой функциональных требований и к практике каталогизации, и к специфическим разработкам Виртуальной библиотеки «Полимат». Проект нацелен на сбор библиографической и иной информации о научной и творческой испано-американской мысли – с исторической точки зрения и на установление связей с другими «цивилизациями» (если использовать понятие Тойнби)⁴.

³ О проекте: Agenjo Bullón, Xavier y Hernández Carrascal, Francisca. *La Biblioteca Virtual Larramendi: fuente de información bibliográfica para el pensamiento iberoamericano en la Web 3.0. --: Jornadas Virtuales Iberoamericanas de Bibliotecología*. [<http://goo.gl/VbwtN>]. Эти страницы написаны осенью 2010 г., после чего были внедрены некоторые важные изменения, о некоторых из них можно прочесть: Agenjo, Xavier y Hernández, Francisca. *La Biblioteca Virtual Ignacio Larramendi desde la perspectiva LOD y EDM – доклад на семинаре I Seminario Internacional de la Biblioteca de Galicia* [<http://goo.gl/uFrXx>].

⁴ Тойнби в «Записках об истории» выделяет 21.

Источники информации для АЗ

Понятно, что многие источники информации, полезные с точки зрения целей библиографического проекта, недоступны в оцифрованном виде, а если и были оцифрованы, то им все равно не хватает соответствующей структуры. В иных случаях, когда ресурсы доступны в веб-среде, дизайн интерфейса уменьшает информационную эффективность. Например, *Diccionario Biográfico Español de la Real Academia de la Historia* даёт капитальную информацию о тысячах лиц, связанных с испанской историей и культурой (в широком смысле), однако технологические возможности этого словаря очень слабы. Другой источник информации - *Hombres y documentos de la filosofía española* (автор -Gonzalo Díaz y Díaz) в семи томах, вышедших с 1980 по 2003, который грешит избыточной эрудицией и недостатком эвристического подхода, но даёт доступ к огромному объёму информации⁵. То же можно сказать и о великольном *Diccionario de filosofía of Ferrater Mora*⁶ - гораздо более строго выстроенном с позиций интеллектуала, но и более скудном – с точки зрения информации.

Очень жаль, что библиографический репертуар, в своё время напечатанный на бумаге и все еще защищенный авторским правом, не доступен в веб-среде и, в частности, недоступен для таких технологий, как Linked Open Data. Парадоксально, но огромные проекты ретроспективной оцифровки останавливаются (в зависимости от страны и соответствующего законодательства) примерно через 50, 70 или даже 80 лет после смерти автора, и поэтому устаревший контент абсолютно доступен, а другой – более современный и поэтому более ценный – недоступен. Говоря о проблемах открытого доступа, нередко забывают о том, что дело не в том, что электронная информация доступна или не открыта. Дело в том, что этой информации в оцифрованном виде просто нет. Последнее обстоятельство (и это необходимо принять во внимание на конференции библиотекарей, для которых библиография – больше чем для кого-нибудь другого - должна быть фундаментальной дисциплиной) заставляет нас забыть о том, что часто лучшие критические издания с приложенными рукописями и набросками (черновиками) тоже недоступны на Вебе. Разумеется, все чаще попадают и восхитительные проекты, где техника электронных изданий применяется к историческим текстам, но все-таки они в меньшинстве. В качестве примера беспрецедентного качества мы приводим *Mark Twain Papers & Project*⁷ - электронное издание значимого автора.

К счастью, Виртуальная библиотека «Полилат» пыталась (по меньшей мере) учитывать мнение учёных – с целью представить обзор состояния библиографических исследований каждого конкретного автора (полимата). Более того - был выявлен минимум ресурсов, по которым следовало наводить справки во всех случаях: *Diccionario Biográfico de la Real Academia de la Historia*⁸, *Fichero de Autoridades de la Biblioteca Nacional*⁹, *Virtual International Authority File*¹⁰ (сюда вошли авторитетные

⁵ Доступно в *Biblioteca Saavedra Fajardo de Pensamiento Político Hispánico* [<http://goo.gl/NHiVW>].

⁶ Ferrater, Mora J. *Diccionario de Filosofía*. Madrid: Alianza, 1979. Это – определённое издание Ferrater Mora, после чего появилось расширенное издание под ред. профессора Jose Maria Terricabras.

⁷ <http://bancroft.berkeley.edu/MTP/>

⁸ <http://www.rah.es/diccBiografico.htm>

⁹ <http://catalogo.bne.es/uhtbin/authoritybrowse.cgi>

файлы национальных библиотек Испании и Португалии), *Library of Congress Authorities*¹¹, и упомянуто печатное *Hombres y documentos de la Filosofía Española y Diccionario de Filosofía*. Логично употребление таких источников как *Stanford Encyclopedia of Philosophy*¹², *The Catholic Encyclopedia*¹³, *Enciclopedia católica*¹⁴, *Jewish virtual Library*¹⁵, *Islamic philosophy Online*¹⁶, *Proyecto Filosofía en Español*¹⁷, *Biblioteca Miralles*¹⁸, *Biblioteca Saavedra y Fajardo*¹⁹, *Biblioteca Virtual Miguel de Cervantes*²⁰, *Proyecto Sarmiento*²¹ и других.

От авторитетных записей к библиографическим

Виртуальный международный авторитетный файл был включён в работу первым. Именно в нем были найдены авторы, отраженные в таких авторитетных источниках информации, как Библиотека Конгресса, Британская библиотека или Немецкая библиотека. Огромное преимущество VIAF состоит в наличии возможности проводить навигацию от кластеров авторитетных данных к каждому из составляющих их заголовков, происходящих из разных авторитетных файлов – поставщиков VIAF. Это позволяет расставлять приоритеты при выборе по источникам: для испанских авторов – НБ Испании, для португальских – НБ Португалии, для бразильских – НБ Бразилии²² и т.д. И для всех – Библиотека Конгресса. Исключительно важно то, что по запросу можно выстроить новую, очень важную и универсальную функцию, заключающуюся в построении электронного агрегата всех библиографических записей, связанных с авторитетными записями. Таким образом, имя автора и его варианты могут быть расположены в одном месте, а это может повлечь за собой значительное обогащение – за счет связанных ассоциациями библиографических записей, которые – по большей части – выполнены национальными библиографирующими учреждениями. В настоящее время VIAF может соединиться почти с двадцатью авторитетными файлами, через которые он даёт доступ к большому числу библиотечных каталогов. Важно отметить, что поля 100 ого блока лишь в некоторых случаях имеют связь с

¹⁰ <http://viaf.org/>

¹¹ <http://authorities.loc.gov/>

¹² <http://plato.stanford.edu/>

¹³ <http://www.newadvent.org/cathen/>

¹⁴ <http://ec.aciprensa.com/>

¹⁵ <http://www.jewishvirtuallibrary.org/>

¹⁶ <http://www.muslimphilosophy.com/>

¹⁷ <http://www.filosofia.org>

¹⁸ <http://www.bibliotecamiralles.org/escritores.html>

¹⁹ <http://saavedrafajardo.um.es/Biblioteca/IndicesW.nsf/Inicio?OpenForm&m=2>

²⁰ www.cervantesvirtual.com

²¹ <http://www.proyectosarmiento.com.ar/>

²² Не вошло в VIAF

полями блогового блока библиографических записей. Если бы это присутствовало везде, то был бы обеспечен доступ и к произведениям об авторе²³, и тогда можно было бы использовать VIAF как хаб - распределительный узел библиотечных ресурсов, не только об авторе, но и о произведениях этого автора.

Источники информации для библиографических записей

Те же механизмы, которые были применены для определения числа источников для авторов, были применены для установления библиографических источников. Таким образом были выбраны: *Catálogo Colectivo de Patrimonio Bibliográfico*²⁴, *Heritage of the Printed Book Database* (HPB) Консорциума научных библиотек Европы CERL²⁵, *Novum Regestrum*²⁶ и *WorldCat*. Особо надо упомянуть Menéndez Pelayo Virtual Library²⁷. Эта виртуальная библиотека продолжает библиографическую работу и работу по составлению, начатую Мендесом Пелайо в работе «*La ciencia española*»²⁸ посвященной ведущим испанским мыслителям и их влиянию, созданную в ответ на знаменитый вопрос Николая Массона²⁹: «А что же Испания? Что сделала она для Европы – двумя, четырьмя, десятью веками позже?»

Ещё мы должны упомянуть два крупномасштабных проекта, проводимых параллельно во времени, но с иным размахом: они затрагивают Виртуальную библиотеку «Полимаг» не только в части контента, но и с точки зрения функциональных требований. Это проекты Испана (*Hispana*³⁰) и Европеана (*Europeana*³¹). Что касается *Hispana*, - есть лишь несколько сообщений, в подробностях описывающих её создание и эволюцию, и это удивительно, учитывая характеристики и объемы проекта, хотя сделано множество презентаций. В момент, когда evolution, which is surprising given the characteristics and dimensions of the project, although there are numerous presentations³². На 30.04.2011 г.

²³ Особенно интересен доступ через VIAF в Библиотеки и архивы Канады [<http://www.collectionscanada.gc.ca/>].

²⁴ <http://www.mcu.es/bibliotecas/MC/CCPB/index.html>

²⁵ Heritage of the Printed Book Database (HPB) del CERL. <http://www.cerl.org/web/en/resources/hpb/main>

²⁶ <http://goo.gl/fqtHA>. См. также *Novum Regestrum: el Catálogo Colectivo del Patrimonio Bibliográfico Iberoamericano* de Xavier Agenjo Bullón, Francisca Hernández Carrascal, *Boletín de la ANABAD*, ISSN 0210-4164, Tomo 44, Nº 4, 1994, 127-142 <http://dialnet.unirioja.es/servlet/articulo?codigo=50938>

²⁷ <http://www.larramendi.es/i18n/bvmpelayo/inicio.cmd>

²⁸ Menéndez Pelayo, Marcelino. *La ciencia española*. In: *Biblioteca Virtual Menéndez Pelayo*. <http://goo.gl/iB5m>

²⁹ Nicolas Masson de Morvilliers (1740-1789)

³⁰ <http://hispana.mcu.es>

³¹ <http://europeana.eu>

³² *Hispana y las iniciativas del Ministerio de Cultura* María Antonio Carrato Mena *Jornada de Difusión de EuropeanaLocal*, 17 nov. 2010 [<http://hdl.handle.net/10421/4765>]. *La aplicación del Modelo de Datos de Europeana a la Biblioteca Virtual de Patrimonio Bibliográfico: bvpb.mcu.es*, María Luisa Martínez-Conde *Jornada de Difusión de EuropeanaLocal*. 17 nov 2010 URI: <http://hdl.handle.net/10421/4783>. More recent: Carrato, María Antonia. *Hispana. I Seminario Internacional de la Biblioteca de Galicia* [<http://goo.gl/6xIqq>]

Hispana имеет 3,181,786 электронных объектов из 146 хранилищ Испании. Очевидно, что материалов значительно больше, но для данного выступления важно упомянуть новую модель данных *Definition of the Europeana Data Model Elements*, версия 5.2.1³³, обновлена 07.03 2011 г., и *Functional specification for Europeana the Danube release*, опубликована 31.08. 2010³⁴.

Не случайно и то, что разработки, выполненные для Виртуальной библиотеки «Полимат» и для фонда Игнасио Ларрамеди так сильно ассоциируются с функциональными спецификациями Europeana. Причина заключается в следующем: участие Испании в проекте Europeana происходит почти исключительно через Испану (Испана поставила 1,367,808 оцифрованных объектов, далеко опередив виртуальную библиотеку Сервантеса = *Biblioteca Virtual Cervantes*³⁵, которая поставила всего 19,062, хотя именно о ней говорят значительно больше). Получается, что и Испана, и большинство провайдеров контента, собранного ею³⁶ работают в ILS DIGIBIB, сейчас – по версии 6.0, но к тому моменту, когда мы будем выступать с докладом в Пуэрто-Рико, уже перейдут на версию 7.0.

Не следует упускать из виду и то, что компания DIGIBIS³⁷, которая разрабатывает ILS DIGIBIB, а также DIGIARCH – для архивов и хранилище OAI под названием OAsIs, это компания, принадлежащая фонду Игнасио Ларрамеди, а Виртуальная библиотека «Полимат» систематически используется в качестве опытного участка для их разработок.

И вот поэтому (и по желанию клиентов) многие описанные здесь средства затрагивают примерно тридцать крупнейших электронных библиотек Испании, и всего – более ста библиотек, если считать библиотеками коллекции, которые совместно собрали «*Biblioteca Virtual de Patrimonio Bibliográfico*» и «*Biblioteca Virtual de Prensa Histórica*».

³⁸

³³ <http://goo.gl/ojLL>

³⁴ <http://goo.gl/P6jme>

³⁵ <http://www.cervantesvirtual.com/>

³⁶ Biblioteca Virtual de Prensa Histórica, Galiciana: Biblioteca Digital de Galicia, Biblioteca Virtual de Andalucía, Gredos (Universidad de Salamanca, Spain), Biblioteca Digital de Madrid, Biblioteca Digital de Castilla-La Mancha, Centro de Documentación de Fundación MAPFRE, Biblioteca Digital de Castilla y León, Biblioteca Virtual del Patrimonio Bibliográfico, Biblioteca Virtual de Derecho Aragonés, Biblioteca Digital Real Academia de la Historia, Catálogo Colectivo de la Red de Bibliotecas de los Archivos Estatales, Biblioteca Virtual del Principado de Asturias, Archivo de la Imagen de Castilla La Mancha, Biblioteca Valenciana Digital, Biblioteca Virtual de Aragón, Fundación Sancho el Sabio, Biblioteca Regional de Murcia, Biblioteca Digital de Aranjuez, Universidad de La Laguna, Biblioteca Virtual de la Diputación de Zaragoza, Biblioteca Virtual de La Rioja, Fundación Ignacio Larramendi, Biblioteca Virtual de la Real Academia Nacional de Farmacia.

³⁷ <http://www.digibis.com>

³⁸ Можно увидеть в Biblioteca Virtual de Patrimonio Bibliográfico – субдомен: *Iberoamérica en las colecciones de la BVPB* [<http://goo.gl/8Jyfa>]. Целесообразно также искать по всем полям в Biblioteca Virtual de Prensa Histórica [<http://prensahistorica.mcu.es>].

Когда мы будем выступать с этим докладом в Пуэрто-Рико, уже будет запущена очередная фаза Европейаны – «Дунай», запланированная на 1 мая 2011 года. Разумеется, мы не можем быть полностью уверены, что именно так и будет, но – так или иначе – некоторая задержка происходит независимо от того, что было запланировано виртуальной библиотекой «Полимат», поскольку новые семантические средства Европейаны действительно оказали решающее воздействие на разработку новой семантической структуры данных. Что касается этих средств, то и Европейана, и Испана – это первичные источники информации виртуальной библиотеки «Полимат» и два методологических примера, с одним из которых (Испана) имеется обратная связь.

На сегодня Европейана выпустила API³⁹ на основе OpenSearch, который можно интегрировать в поиск по любому веб-сайту и в показ записей, найденных в Европейане. Такой API, работающий очень эффективно и уже инкорпорированный в веб-интерфейс виртуальной библиотеки «Полимат», автоматически запускает поиск по изначальному запросу в базе данных Европейаны. Т.е., запрос, сделанный в базу данных, состоящую из нескольких тысяч записей, проходит и через базу данных Европейаны, в которой около двадцати миллионов электронных объектов (17,901,019).

2. Электронные агрегаты

Согласно устоявшейся методике, ядерные компоненты Виртуальной библиотеки «Полимат» называются «электронными или цифровыми агрегатами», которые даже по стилю очень близки к статьям традиционных библиотек, при этом они пытаются достичь уровня сложности и полноты некоторых статей Википедии (или DB-педия, описываемая ниже).

Термины «агрегат» и «агрегация» употребляются по-разному, - в зависимости от среды, в которой они работают, и могут повергнуть читателя в некоторую неопределённость. И вот, в процессе сбора метаданных те сборщики, которые сами могут быть собраны, называются агрегаторами – как в Европейане. Однако, понятие электронного или цифрового агрегата Виртуальной библиотеки «Полимат» более схоже с определением, которое сделано Движением открытого архива для повторного использования объектов и обмена ими (Open Archives Initiative Object Reuse and Exchange = OAI-ORE)⁴⁰, где агрегат – это ресурс, в котором содержится набор взаимосвязанных ресурсов. Тогда, электронная агрегация – это набор данных и электронных информационных ресурсов, собранных «вокруг» конкретного автора, ядром которого является авторитетная запись.

Процесс авторитетного контроля, выполняемый фондом Игнасио Ларраменди, объемнее того, который традиционно имеет место в крупных библиотеках или библиографирующих учреждениях. Мы настаиваем не на выполнении функции идентификации сущности «лицо», а на выполнении задач текстуализации, - тем способом, который определён FRAD.⁴¹ Таким образом, не только вносятся необычные

³⁹ <http://www.version1.europeana.eu/web/api>

⁴⁰ <http://www.openarchives.org/ore/>

⁴¹ [21](#)

данные в АФ, но и устанавливается связь от записей к другим ресурсам авторитетных данных, и иным источникам и данным, доступным в веб – среде. В этом случае источники полученных данных (они записываются в поле 670) являются не только оправданием выбора имён и их вариантов, но и отслеживают источники библиографической информации о конкретном выбранном авторе.

Подобно тому, как это делается в VIAF, каждая авторитетная запись получает определение, согласно формату MARC 21, для основного заглавия (поля 1ого блока), вариантов имён (поля 4ого блока), связи с другими именами (поля 5ого блока) и для эквивалентных заглавий на других языках (поля 7ого блока). Последнее особенно важно для виртуальной библиотеки «Полимат», поскольку авторы отобраны не только из-за их значения для испано-американской культуры, но из-за вклада испано-американской культуры во всемирную культуру, и вот поэтому иметь эквиваленты на других языках абсолютно необходимо. Именно так возможно следить за распространением текстов по всему миру, а также – за распространением влияния идей. Та же ситуация имеет место и в Испании, где наряду с испанским (кастильским) языком, существуют и другие: каталанский, валенсийский, галисийский и баскский.

Многоязычные заголовки

Нам хотелось бы обратить внимание аудитории на необходимость наличия в формате MARC 21 специфического элемента для кодировки языка заголовков. К сожалению, создается впечатление, что обсуждение многоязычных авторитетных записей в формате для представления авторитетных данных (Discussion Paper 2001-DP05 Multilingual Authority Records in the MARC 21 Authority Format⁴²), не закончилось каким бы то ни было предложением, по крайней мере – на нынешний день. Поле 7ого блока позволяют записывать эквивалентные заглавия на разных языках, но не имеют средства для обозначения языка заголовка, что особенно необходимо в многоязычных авторитетных файлах. Другие механизмы также имеют ограничения: позиция 008/08 кодирует только английский и/или французский языки каталога, а подполе \$ 040 b позволяет только установить язык каталогизации, который часто отличается от языка заголовка.

Эту ситуацию разумно преодолели в MADS и, разумеется, в предложенном MADS / RDF⁴³, но не следует забывать, что большинство авторитетных записей существует только в MARC формате и поэтому будет трудно немедленно переместить данные этого типа в среду связанных открытых данных (Linked Open Data environment). Так что VIAF унаследовал эту проблему, и, если нет возможности опознать учреждение и страну происхождения записи, то язык, на котором создан заголовок, можно определить только методом дедукции. Это особенно важно в тех случаях, когда заголовки созданы в одной лингвистической области, и могут быть очень похожи, а то и неразличимы (например, *Geografía* на испанском *Geografia* [без акцентов] на итальянском или каталанском языках).

⁴² <http://www.loc.gov/marc/marbi/2001/2001-dp05.html>

⁴³ MADS/RDF Primer. Final Public Review Document. [<http://www.loc.gov/standards/mads/rdf/>]

Варианты имен лиц

Поиск информации в других авторитетных файлах проводится не только для установления отношений лингвистической эквивалентности между заголовками, но и для увеличения количества вариантов этих имён и связей с другими именами. Эта процедура, соответствующая поведению ILS DIGIBIB, позволяет импортировать библиографические записи, должным образом связанные с установленным заголовком, а также – контролирует создание не установленных записей в полях 4ого блока. Процесс проходит несколько медленнее в начале – в ходе создания авторитетной записи, но он гораздо продуктивнее в середине срока и качество гораздо выше. Необходимо также заметить, что оригинальный библиографический проект будет иметь лучшее завершение, если источники анализировались, начиная с ядра авторитетных записей.

Эта процедура особенно важна, когда мы видим группу авторов, от Сенеки (Кордова, 4 до н.э. - Рим, 65 н.э.) до Мартина де Рикера (Барселона, 1914 -) и многие из них пользовались разными именами, в зависимости от языков, на которых они писали, не говоря уже о том, что эти имена были переведены на многие языки вместе с произведениями (например, Avergoes – это латинское имя Ibn Rus). Чтобы иметь большое количество вариантов имён, адекватных ресурсам, которые предстоит импортировать, необходимо располагать достаточно широким спектром наименований, которые позволят располагать нужным количеством позиций для импорта, но не более чем это необходимо, т.к. они уже зарегистрированы в государственных органах - во многих версиях.

3. Авторитетные записи

Обновление No. 12 для формата MARC 21, опубликованное в октябре 2010 года, консолидирует множество полей, следуя философии RDA, что существенно для агрегации информации о лицах. Конечно, речь идёт о полях 3го блока. По ходу дела элементы, которые изначально предназначались для того, чтобы отличать друг от друга омонимы, постепенно приобрели собственную значимость. Так, различные подполя поля 100, которые рекомендовалось использовать только для того, чтобы отличать авторов с одинаковыми именами, стали обязательными, в особенности – поле \$d, немедленно контекстуализирующее автора с исторической точки зрения, что неоднократно помогало взаимодействовать с произведением и избегать ошибок при атрибуции. До того, как RDA были приняты MARC 21⁴⁴, не было атрибутов личности, более выразительных, чем место рождения. Именно эти атрибуты относятся не к *имени* лица, а к самому лицу, и помогают придать записям MARC-формата энциклопедический вид, почти db-педический, если мы можем использовать такой неологизм.

Выход RDA оказался исключительно полезным для определения направлений работы виртуальной библиотеки «Полимат». Во-первых, при поиске информации об авторах, согласно пользовательским задачам, определённым в FRAD, как поиск и контекстуализация. Во-вторых, структуризация своих информационных ресурсов по форме, смоделированной в RDA: Произведение – Выражение – Воплощение –

⁴⁴ <http://www.loc.gov/marc/RDAinMARC29.html>

Физическая единица⁴⁵. По этой причине в следующие версии DIGIBIB немедленно внедрялись обновления MARC 21, в той мере, в которой ILS DIGIBIB соотносило свои новые выпуски с обновлениями MARC 21.

Насколько это возможно, для каждого автора помимо канонической формы имени включены варианты имени и его формы на других языках, биография автора (678), даты рождения и смерти (046), места рождения и смерти (370) и другие места, связанные с деятельностью (370), род занятий (372 и 374), язык, которым он пользовался (377) и его пол (375). Эти атрибуты личности ставят окончательной целью поддержку семантического поиска и навигации между данными. Это позволяет веб-сайту виртуальной библиотеки «Полимат» предлагать пользователям возможность искать авторов по годам, когда они активно творили, по местам, связанным с их биографиями, по профессиям и роду занятий, - т.е. не только по их произведениям, что одинаково для всех ОРАСов.

Авторитетные записи в других схемах метаданных

Виртуальная библиотека «Полимат» в равной степени заинтересована в информации об организациях, лицах и семействах (родах), находящейся в архивах. Необходимость конвергенции библиотечных, архивных и музейных данных давно подтверждена, и имеются примеры того, что такую работу можно провести на высоком уровне. Эти примеры чаще всего появляются как интерфейсы для объединенного поиска, и это способствует усилению информационных возможностей таких данных. Совсем недавно DIGIBIS провел работу такого рода с проектом Правительства Каситы-Ла-Манчи, посвященным двум архивам: Archivo de la Imagen⁴⁶ (фотографии) и Archivo de la palabra (звукозаписи)⁴⁷. Для этого производилось систематическое сопоставление⁴⁸ MARC 21 с EAD и EAC-CPF, а также – с ISAD (G) и ISAAR (CPF).

Схожим образом, как и ожидалось, виртуальная библиотека «Полимат» уже потребовала возможности установить связь с библиографическим и архивными данными, относящимися, например, к Агустину де Монтаниано и Луяндо (Agustín de Montiano y Luyando, 1697-1794), историку, критику, автору пьес, вошедшему в историю в качестве первого директора Королевской исторической академии, основанной в 1735 году, - поскольку его документы являются частью Виртуальных библиотек Ларраменди⁴⁹.

⁴⁵ MARC Discussion Paper No. 2011-DP03: Identifying Work, Expression, and Manifestation records in the MARC 21 Bibliographic, Authority, and Holdings Formats [<http://www.loc.gov/marc/marbi/2011/2011-dp03.html>]

⁴⁶ http://clip.jccm.es/archivo_de_la_imagen/es/micrositios/inicio.cmd

⁴⁷ http://clip.jccm.es/archivo_de_la_palabra/es/micrositios/inicio.cmd

⁴⁸ http://www.sedic.es/p_boletin_novedades_mensual.asp#Digibis

⁴⁹ Это предок Ignacio de Larramendi y Montiano (1921-2001) основателя фонда, поддерживающего, в частности Larramendi Virtual Libraries или DIGIBIS. См. также: Hernando de Larramendi, Ignacio. *Mecenazgo cultural de Ignacio Hernando de Larramendi y Montiano : crónica y testimonios*. [<http://www.larramendi.es/i18n/consulta/registro.cmd?id=954>]

4. Связывание данных

Виртуальная библиотека «Полимаат» является провайдером контента для Европеаны, собранной Испанией, которая, в свою очередь является национальным агрегатором в рамках Европеаны (как уже было упомянуто) и применяет её базу данных как источник информации. В связи с этим Виртуальная библиотека «Полимаат» поставила стратегическую задачу: принять модель, предложенную Моделью данных Европеаны (EDM)⁵⁰ и технологиями Связанных открытых данных (Linked Open Data). И EDM, и Linked Open Data (LOD) очень быстро подходят к новой деятельности и поставляют огромные объёмы данных – как в случаях с VIAF⁵¹ или DBpedia⁵². С того момента, как Тим Бернерс-Ли в 2006 году предложил Связанные данные, и до настоящего времени семантически структурированные данные, пригодные для многократного использования, публиковались в количествах, прежде не встречавшихся в веб-среде; возможно, единственным исключением были библиографические записи в library MARC формате. Говоря о вкладе библиотек в LOD, следует упомянуть, что в 2010 году в рамках W3C была создана Группа инкубатора библиотеки связанных данных (LLD)⁵³. Эта группа отвечает за подготовку нескольких отчётов и докладов о статусе онтологий для архивов, библиотек и музеев; словарей и подборок данных, представляющих интерес для использования в этих областях и в тех случаях, которыми занимается библиотечное сообщество. Чтобы все заинтересованные лица могли следить за деятельностью группы и Linking Open Data⁵⁴, информирование будет налажено соответствующим образом.

Для Виртуальной библиотеки «Полимаат» Linked Open Data⁵⁵ и прилагаемые технологии представляют собой возможность полной реализации подхода к «электронным агрегатам» как к узлам Связанных данных. Фактически, Виртуальная библиотека «Полимаат» - это случай использования, рассмотренный в LLD. Речь идёт не только о считывании информации, но и о связи её с соответствующими узлами LOD и его ресурсами согласно правилам игры, где от каждого ожидают исполнения его роли. Каждый ресурс, доступный в том или ином наборе данных LOD, идентифицируется URI со специальными характеристиками⁵⁶. Этот URI тащит за собою сеть отношений, связей, атрибутов к другим ресурсам, которые, согласно используемой модели (онтология или Схема RDF), определяют сущности и принадлежность данной функциональной области. Их значимость опирается на то, что они позволяют выявить, что некая сущность из одной среды точно такая же или очень

⁵⁰ Doerr, Martin, Gradmann, Stefan, Henniecke, Steffen, ISAAC, Antoine, Meghini, Carlo, and SOMPLE, Herbert van de. *The Europeana Data Model (EDM)*. En: 76th IFLA General Conference and Assembly, 10-15 August, 2010, Gothenburg, Sweden. [<http://goo.gl/7eLs>]

⁵¹ <http://ckan.net/package/viaf>

⁵² <http://dbpedia.org/About>

⁵³ <http://www.w3.org/2005/Incubator/ld/>

⁵⁴ <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁵⁵ http://www.w3.org/2005/Incubator/ld/wiki/Use_Case_Polymath_Virtual_Library

⁵⁶ <http://www.w3.org/TR/2007/WD-cooluris-20071217/>

похожа на другую сущность, выявленную в другой среде. Так, для DB-педии Маймонид – это философ, а в VIAF – имя лица. Процесс состоит из связывания таких данных через URI, открытия их для повторного использования и, анализа их с помощью определённого инструментария⁵⁷, выявления других данных из их связей и правил проникновения, а также – обновления оригинальных данных так, чтобы использовать их снова после модификации и расширения. И так далее – после каждой итерации.

DB-педия и VIAF

В 2010 году Виртуальная библиотека «Полимат» начала процесс установления связей авторитетных записей с похожими ресурсами, существующими в LOD. Очевидно, что источники информации об изложенной здесь методологии стали кандидатами на установление связи. Итак, URI ресурсов VIAF регистрируют в 024 поле записей формата MARC21. Ресурс RDF **Maimonides** (rdf: about = "viaf/100185495"), группирует установленные заголовки для этого автора из семи различных авторитетных файлов с 255 вариантами имён, а также имеет связи к контенту DB-педии (<http://dbpedia.org/resource/Maimonides>). Если опытный библиотекарь уже работал в этой области (например, главный каталогизатор), он может извлечь из сети связей VIAF до 12 различных заголовков для Маймонида⁵⁸, из разных стран и на разных языках. Ясно, что процесс получения этих данных вручную очень точен, но замедлен, и скоро стало очевидно удобство автоматизации процесса, - после того, как преимущества и недостатки были взвешены. Несомненно: число авторов, составляющих Виртуальную библиотеку «Полимат» (около тысячи) можно было бы обработать вручную, но продуктивность с точки зрения времени и данных можно было повысить путем автоматизации. К тому же, раз инструменты были созданы и разработаны, их можно было инкорпорировать в DIGIBIB. Ожидается, что на момент представления этого доклада 13-18 августа 2011 года эти инструменты будут составлять модуль DIGIBIB 7.0 и следующих версий.

SKOS: другой путь к Linked Open Data

Эти типы связей устанавливались и с применением SKOS. Результаты показывают, что частное понятие – то же или очень похоже на другие понятия в других словарях. Через предметные рубрики библиографических записей, связывая понятия с другими словарями, тоже доступными в SKOS, также можно получить достаточное число произведений об авторе. Процесс поиска информационных ресурсов может происходить автоматически, охватывая сеть связей между понятиями из различных словарей, при наличии возможности обнаруживать информационные ресурсы, связанные с этими понятиями.

В настоящее время Виртуальная библиотека «Полимат» работает (в тестовом режиме) с Списком предметных рубрик для публичных библиотек Испании, составленным Генеральным управлением книг и библиотек, и расширенным за счет «скрещивания» с

⁵⁷ <http://www.w3.org/wiki/SemanticWebTools>

⁵⁸ Ibn Maṣmun, Musà или, более привычное для ортодоксальных иудеев имя – Rambam.

авторитетным файлом НБ Испании⁵⁹, чтобы сделать его эквивалентным Списку предметных рубрик Библиотеки Конгресса⁶⁰. Поскольку данный процесс аналогичен процессу по установлению связей между понятиями SKOS из RAMEAU⁶¹ с Предметными рубриками Библиотеки Конгресса и Предметными рубриками Немецкой библиотеки (SWD)⁶², было бы разумно установить такие концептуальные связи, которые определённно не будут обладать большой гранулярностью, но, обработанные и скрещенные с определённым числом авторов, могли бы предложить понятную связь между крупными совокупностями информации.

Время и пространство

Еще два элемента – хронологические и географические данные – играют важнейшую роль в агрегации и связывании данных, имеющих отношение к любому автору. Действительно: любой автор живёт и умирает в определённое время, он работает, поддерживает отношения с современниками в течение разных периодов, пишет книги, которые издаются, переводят, редактируют в соответствующие годы⁶³. Возможность связывать данные авторов этой новой хронологической категорией, а не только именем или предметом, уточняет совокупность произведений автора, хотя понятно, что будет заметна непоследовательность или прерывистость. Кодировка времени, присутствующая во многих элементах авторитетных и библиографических записей, нередко нуждается в атрибутах для выражения предположений, неточностей, оценочных суждений и т.д., так часто встречающихся в описаниях людей и ресурсов, не говоря уже о событиях.

Такие данные важны для составления линии жизни или для выполнения операций и извлечения данных из дат, как это можно видеть, например, в WorldCat Identities⁶⁴.

Похожая работа протекает в агрегации и связывании географических данных. Фактически, внутри Linked Open Data присутствует возможность для получения данных этого типа – географических имён, GeoNames⁶⁵, которые также представляют собою единственный известный сегодня случай в Испании трансформации географических кодов Национального географического института (Instituto Geográfico Nacional⁶⁶) в RDF и LOD. Опять же, и здесь можно расширить процесс агрегации.

⁵⁹ *Autoridades de la Biblioteca Nacional* [Recurso electrónico]/ Biblioteca Nacional ; software, Chadwyck Healey -- Número 1 (nov. 1996). - Madrid : Biblioteca Nacional : Chadwyck-Healey España, 1996- CD-ROM.

⁶⁰ <http://id.loc.gov>

⁶¹ <http://www.cs.vu.nl/STITCH/rameau/>

⁶² <http://www.d-nb.de/standardisierung/normdateien/swd.htm>

⁶³ There is now a new format that will attempt to standardize the many different ways in which different chronologies have been established and even numbers with those that have been represented, if not through the alphabet, the latin alphabet, the arabic alphabet, etc. Extended Time/Date Format ETDF [<http://www.loc.gov/standards/datetime/>]

⁶⁴ <http://www.worldcat.org/identities/>

⁶⁵ <http://www.geonames.org/ontology/documentation.html>

⁶⁶ GeoLinkedData был запущен вместе с опубликованием различных источников, поступивших от Instituto Geográfico Nacional, что сделало эти источники доступными в качестве основ знаний RDF,

Авторы родились и умерли в конкретных местах, были связаны с современниками в конкретных местах (возможно – школах, университетах, религиозных орденах или военных, научных либо культурных сообществах, а также – политических группах, расположенных в конкретном месте, или связанных с таким местом. Поэтому через них можно установить новые и уместные агрегации и связи. Процесс установления связей между GeoLinkedData и данными GeoNames выполнен автоматически, путем конвертирования отобранных данных в структуру, которую сможет усвоить DIGIBIB. Чтобы автоматически все подполя 752 и 151 поля, эти данные были сопоставлены с MARCcom. Таким образом, мы получили не только имена, имеющиеся в этих словарях, но и географические координаты. Несомненно, далее должна быть проведена работа вручную, чтобы разрешить проблемы, которые могут возникнуть, и «развести» географические понятия и административные, а также географические места (местонахождения) и устаревшие административные. Ниже описывается разница между тем, как работать с *лицами*, и с *именами лиц*. Ясно, что управление географическими наименованиями, наименованиями административных делений и географическими расположениями предполагает различную структуру данных.

5. Использование LOD

Можно сделать запрос на веб сайте VIAF и полученные результаты могут быть выведены на показ в MARC XML или RDF. Можно проводить поиск в VIAF через его API⁶⁷ - по SRU и OpenSearch, а результаты можно выгрузить в MARC XML, RDF и разные другие форматы и схемы. VIAF⁶⁸ доступен и как Linked Open Data. Исходя из рекомендаций показывать данные в LOD⁶⁹, зачастую получить доступ к набору данных и сделать запрос необходимо с помощью клиентов SPARQL Endpoint⁷⁰ или SPARQL⁷¹, которые можно использовать для запросов в файлах RDF. Значит, можно проводить поиск по DB-педии⁷² или VIAF через SPARQL⁷³ и тогда запросы могут логически расширяться относительно данных, т.е. можно найти гораздо больше связанной информации, чем при способах поиска, описанных выше. Используя SPARQL, можно составить более сложное поисковое предписание: пример из DB-педии, но не из VIAF, это 'Философы, на которых оказал воздействие Маймонид'. И вот, виртуальная библиотека «Полимат» делает новые шаги – на сей раз по

согласно принципам Связанных данных (Linked Data)
<http://geo.linkeddata.es/web/guest;jsessionid=6E6A9C1E1DAF40F81005E8F4DA3A27E1>

⁶⁷ <http://www.oclc.org/developer/documentation/virtual-international-authority-file-viaf/using-api>

⁶⁸ <http://ckan.net/package/viaf>

⁶⁹ *How to Publish Linked Data on the Web*. <http://www4.wiwiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial/BestPracticeRecipesforPublishingRDFVocabularies>. W3C Working Group Note 28 August 2008. <http://www.w3.org/TR/swbp-vocab-pub/>

⁷⁰ <http://www.w3.org/wiki/SparqlEndpoints>

⁷¹ <http://www.w3.org/wiki/SparqlImplementations>

⁷² <http://ckan.net/package/dbpedia>

⁷³ SPARQL Query Language for RDF. [<http://www.w3.org/TR/rdf-sparql-query/>]. Сейчас W3C работает на SPARQL 1.1 Federated Query. [<http://www.w3.org/2009/sparql/docs/fed/service>].

библиографическому плану, чтобы воспользоваться кое-какой библиографической информацией, «просачивающейся» в виде наборов данных LOD. И делает это как с точки зрения потребителя, так и с точки зрения производителя.

Первый шаг – это внедрение Европейской модели данных (Europeana Data Model = EDM) в ответ на изменения, проведение которых запланировано Европеаной в ходе фазы «Дунай». У Европеаны есть возможность более-менее автоматически оценивать данные для того, чтобы переводить в новые EDM информацию, полученную от провайдеров контента, но, учитывая, что этих провайдеров уже несколько тысяч, работающих на двадцати пяти языках и с таким же количеством словарей, мы понимаем, что эти оценки затронут лишь неопределённый процент данных, принесённый результат, который будет невозможно выверить, и, конечно же, очень велик риск установления ошибочных отношений эквивалентности и недействительных связей. Независимо от этого такая процедура может быть практичной и способной разрешить вопрос об установлении связей между данными одной части, не распространяя эти связи на весь объём, расширение может произойти без вмешательства самих провайдеров контента. В типичном случае, географические наименования часто действительно совпадают с названиями административных делений, в отличие от их географических координат, и поэтому сопоставлять Кастилию пятнадцатого века и современную Кастилию – это всего лишь делать некое допущение. Так или иначе, виртуальная библиотека «Полимаг» начала процесс публикации своих данных, имеющих внутренние связи в LOD.

Относительно трудно сравнивать библиографические данные MARC 21 с семантическими элементами Европеаны (Europeana Semantic Elements 3.3.1) и с моделью данных Европеаны (Europeana Data Model). Но, если такой процесс всё же будет иметь место, Европеане будет очень трудно проводить точную агрегацию данных (в духе OAI-ORE) без связывания этих данных с внешними источниками или словарями. Методология инкорпорации URI Виртуального авторитетного файла в авторитетные записи показала себя исключительно полезной, поскольку поколение данных EDM может предложить ссылку на внешний словарь для класса *Агент*.

Второй процесс – установление связей через Список предметных рубрик Библиотеки Конгресса (LCSH), размещённый на *id.loc.gov*. Итак, виртуальная библиотека «Полимаг» может публиковать свои предметные рубрики в SKOS, но они связаны с LCSH. За счёт этого свойство *dc: subject* будет ассоциироваться с *skos:Concept*, связано с Предметными рубриками Библиотеки Конгресса и, опосредовано, с RAMEAU⁷⁴ и SWD⁷⁵ - как минимум. Иными словами, предметные рубрики, доступные в некой семантической структуре и связанные с такими же рубриками, выраженными по-английски, по-французски и по-немецки. Это – ясный пример преимуществ связывания разных наборов данных с одним или несколькими специфическими словарями.

Кроме того, как это ни парадоксально, та же процедура может быть применена для установки взаимных связей между национальными данными. Один из эффектов

⁷⁴ <http://ckan.net/package/stitch-rameau>

⁷⁵ <http://ckan.net/package/dnb-gemeinsame-normdatei>

использования библиотеками Испании Предметных рубрик Библиотеки Конгресса как одного из главных источников для подтверждения предметных рубрик состоит в том, что эта связь (в поле 670) даёт возможность связать различные списки предметных рубрик, которые (и это нельзя забывать!) могут быть на разных языках. Это – исключительно выигрышная сторона эффекта от LOD, - пусть даже опосредованно, это поможет разрешить связывать данные между каталогами библиотек Испании и авторитетными файлами, которые остаются скорее изолированными, чем желанными и понятными.

Здесь мы опять должны указать на другое положительное воздействие Европейцы на информационные системы страны. Министерство культуры Испании, которое владеет Испаной, начало также процесс адаптации Испании к EDM и, как упоминалось, одним из первых шагов стала конвертация Списка предметных рубрик публичных библиотек в (LEM) в SKOS, с установлением связей между записями LEM и Предметными рубриками Библиотеки Конгресса – тех случаях, когда это возможно. Виртуальная библиотека «Полимат» завершила процесс установления связей, связав свои предметные рубрики с LEM, что логично. Высказывается надежда, что через набор данных LEM можно будет установить связь и с другими списками предметных рубрик Испании – на галисийском, каталонском, валенсийском и баскском языках. 30 апреля 2011 года виртуальная библиотека библиографического наследия (Biblioteca Virtual de Patrimonio Bibliográfico) Министерства культуры Испании, внедрило возможность для ILS DIGIBIB 7.0 экспортировать записи согласно EDM 5.2.1.

Эти процессы преобразуются в полное внедрение EDM и в публикацию наборов данных виртуальной библиотеки «Полимат» в LOD. В этом случае публикация данных в виде LOD должна происходить медленно, соответствуя выполнению библиографического плана, чтобы всеобщим достоянием становились только устоявшиеся данные, с разумной степенью доверия и обладающие достаточным количеством связей⁷⁶.

Как библиографический проект виртуальной библиотеки «Полимат» предполагает критический отбор источников, так библиотечный проект публикации в виде LOD подразумевает похожий отбор наборов данных. И не может быть места, подходящего больше, чем Конгресс ИФЛА для того, чтобы вспомнить о том, что деревья схем метаданных, онтологий и приведения онтологий к единообразию не могут заставить нас потерять за ними библиотечный лес, - библиографические задачи наборов данных, которые предполагается опубликовать в LOD

6. Автоматический процесс

Представляя потребителей информации, виртуальная библиотека «Полимат» разрабатывает различные приложения для считывания, интеграции и распространения данных. Мы уже упоминали многократное применение данных из GeoNames или LEM. В о время, когда пишется этот доклад, идут разработки других приложений, для считывания наборов данных LOD с использованием полуавтоматических процедур. Этот процесс может заключаться в считывании некоторых или всех триплетов, связанных с одним или несколькими авторами, или только в получении URI данного ресурса. Эти данные хранятся в промежуточном репозитории, в котором можно

⁷⁶ Weibel, Stuart. *Principles of Linked Data Recast*. Weibel Lines [<http://goo.gl/sYSDe>]

проводить поиск, отбор и обновление операций. Таким образом, сверенные данные могут быть поглощены либо библиографической базой данных, либо авторитетным файлом, при этом имеется возможность выбрать все атрибуты ресурса или только его URI, и обновить записи базы.

То же можно проделать с описаниями из библиографических и информационных ресурсов. Этот процесс должен скоро начаться, но у него имеется дополнительная проблема: правильная идентификация произведений. Если такая идентификация имен лиц в авторитетных файлах унаследовала множество проблем: недостаток данных для контекстуализации, то в случае с произведениями такой недостаток данных может стать ещё большей проблемой. И вот, отсутствие унифицированных заглавий и записей на автора+унифицированное заглавие, или на связи между заглавиями произведений и их транслитерации, делают автоматизированный или полуавтоматизированный процесс особенно нецелесообразным.

Более того: не все источники авторитетных или библиографических данных доступны на LOD, хотя ситуация и меняется ежедневно, - и виртуальная библиотека «Полимаг» использует серверы SRU, на которых можно запускать специфические профили поиска. Этим профилям можно придать такую конфигурацию, чтобы размещать запросы согласно желаемым элементам. Преимущества созданной разработки состоят в том, что результаты могут храниться в репозитории OAI, который регулярно выполняет сбор метаданных с серверов SRU. Мы располагаем большими возможностями, которые может обеспечить эта система для того, чтобы создать национального, а, возможно, и межнационального сборщика. Великолепный прогресс, достигнутый Европейской на сегодняшний день, основан на сборе метаданных OAI и он может послужить началом похожей системы, которую можно было бы назвать Американой. Основным преимуществом этой процедуры стал бы селективный сбор метаданных, и мы надеемся получить хороший результат и на испанских, и на межнациональных ресурсах.

7. Энциклопедия *authorities*

Эта модель переключилась с авторитетных записей для *имен лиц* на записи для *лиц*, - это вовсе не номинальный вопрос, а длинный ряд вопросов. Фактически, некоторые проблемы каталогизации, которые надо решить в ближайшее время, должны будут свести вместе атрибуты *лица* и атрибуты *имени лица*. Онтологии и наборы данных (такие как VIAF или DB-педия) показывают различные дефиниции классов и атрибутов, которые устанавливают имена лиц.

Если в центре облака значений, представленных диаграммой Linked Open Data находится DBpedia, т.е. Wikipedia, закономерно, что представление наших авторитетных записей, напоминающих по форме записи из энциклопедии, напоминало или имело ту же структуру, что у записей Wikipedia, но с использованием в качестве основы MARC-записи. Было бы абсурдом не принимать во внимание огромный объём MARC-записей в мире (тысячи миллионов). Учитывая также подробную детализацию и гибкость MARC 21, его можно легко конвертировать в структуру XML. Конечно, возможно, что любая система может получить данные из любой базы данных, доступной в LOD, но не надо забывать, что дело не только в связывании данных, но и в их ведении и обновлении. В этом смысле, можно организовать структуры DB-педии, VIAF, MADS, MARC21 и EAC-CPF так, чтобы организовать приложение, комбинирующее контент с различных источников, но этого может быть недостаточно,

поскольку само по себе связывание данных может привести к неконтролируемым ошибкам, неточностям или избыточности. Т.е. нам нужна иная структура данных для лиц и имён лиц, которая могла бы поддерживать управление данными и их ведение баз.

Такая структура авторитетных записей может быть комбинацией схем, основанных на MARC формате, анализирующей совместно с MADS / RDF, EAC-CPF⁷⁷ и включающей некоторые атрибуты DB-педии⁷⁸. Вполне возможно, что влияние DB-педии станет необходимым для расширения количества атрибутов, определяющих лицо. Онтология DB-педии демонстрирует некоторые свойства класса «лицо», представляющих действительный интерес для виртуальной библиотеки «Полимаг»; среди прочих это 'Влияние', 'находится под влиянием' или 'философская школа'.

Этот аспект лучше виден в подпроекте (или, лучше – *поднаборе*) виртуальной библиотеки «Полимаг» под названием «Школа Саламанки» (Escuela de Salamanca). Под этим обобщенным именем подразумевается группа испанских и португальских академиков (преимущественно – теологи, юристы и экономисты), которые непосредственно участвовали в возрождении научной мысли в шестнадцатом – начале семнадцатого века, после открытия Нового света, ведущих корни от интеллектуальной и педагогической деятельности Франсиско де Витория в Университете Саламанки. Окончательная цель этого подпроекта – просто проследить влияние этих предшественников в некоторых областях права, политики и – особенно – экономики⁷⁹. Новое время внесло значительные изменения в понятие «человек в обществе», и именно Школа Саламанки по-новому подошла к этим вопросам. Так, Франсиско де Витория, Доминго де Сото, Мартин де Аспилкуэта Томас де Меркадо и Франсиску Суареш делали попытки примирить Томистическую доктрину с новым общественным и экономическим порядком. Так мы обнаруживаем группу авторов, занимавшихся экономической наукой⁸⁰, которые, занявшись моральной легитимностью Конквисты, выработали наиболее новаторские теории⁸¹.

Подводя итог, мы видим, что авторы, вошедшие в виртуальную библиотеку «Полимаг», - авторы нескольких тысяч произведений, которые обнаружены во множестве библиотек при помощи процедуры связывания произведений с другими авторами через VIAF, позже снабжённые виртуальной навигацией между предметами этих книг с использованием словарей или списков предметных рубрик - LCSH, RAMEU или SWD, - эти авторы теперь взаимосвязаны и формируют сеть, которая

⁷⁷ Нам приходится ссылаться в данном сообщении на последнее издание CIDOC-ICOM Linked Open Data Recommendation for Museums CIDOC-ICOM Linked Open Data Recommendation for Museums. [http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html]

⁷⁸ <http://mappings.dbpedia.org/server/ontology/classes/Person>

⁷⁹ Ajenjo Bullón, Xavier; Hernández Carrascal, Francisca y Juez García, Patricia: *La Escuela de Salamanca considerada desde el punto de vista de la Web semántica y la información en la red*. Доклад от 14 апреля 2011 года на конференции: *X Jornadas de la Asociación de Hispanismo Filosófico: Crisis de la modernidad y filosofías ibéricas*, held in Santiago de Compostela, 2011, April, 13-15. [<http://goo.gl/uFrXx>]

⁸⁰ Schumpeter, Joseph A. *History of Economic Analysis*. New York: Oxford University Press, 1954.

⁸¹ Grice-Hutchinson, Marjorie. *The School of Salamanca: Readings in Spanish Monetary Theory, 1544-1605*. Oxford: Clarendon Press, 1952.

прекрасно видна. Эта сеть установила широкую связь через данные, особенно – через метаданные, относящиеся не столько к имени лица, сколько к самому лицу и, без сомнения, к организациям, подход к которым аналогичен подходу к лицам. Теперь мы добавим характерные особенности произведений, т.к в большинстве случаев через свои выражения произведения будут воплощены благодаря полиграфистам, редакторам и книжной торговле (и библиотеками, где их физические единицы хранятся или оцифровываются). И вот теперь все эти сущности также представлены в пространстве и во времени. Таким образом, следом за Школой Саламанки мы увидим, как Конвент Святого Эстебана, в котором расположен университет Саламанки (40° 57' 38" СШ - 5° 39' 47" ВД), играл ключевую роль для студентов и преподавателей, и даже выполнял издательские и редакторские функции, что, впрочем, продолжается и по сей день.

Новое определение функциональности ИБС

Возможности представления связанных данных приведут к полному переосмыслению функций интегрированной библиотечной системы, а именно - DIGIBIB.

1. Возможность инкорпорирования элементов данных или атрибутов, которых нет в формате MARC, и их обновление, управление ими и их экспорт в виде LOD.
2. Возможность поглощения наборов данных – полностью или выбранными фрагментами; это означает возможность обновить всю запись или её часть. Например, возможно, как в виртуальной библиотеке «Полимаст» выбрать и поглотить только URI ресурса.
3. Необходимо модифицировать интерфейсы для запросов так, чтобы они давали возможность находить и получать данные он-лайн из разных ресурсов через API на основе SRU / OpenSearch или даже в сочетании с SPARQL.
4. В конце этого – неполного – списка необходимо иметь возможность управлять другими типами информации, не так широко используемой, в качестве описаний наборов данных LOD в целом, - как пример необходимо упомянуть инструменты архивного поиска и музейные записи.

8. Доступность наборов данных

Следует упомянуть, что помимо доступности наборов, как и всегда, данных важен критический отбор этих наборов данных, которые, в сущности, являются не чем иным как ресурсами библиографической информации с новой структурой. Критический анализ качества данных с интеллектуальной и теоретической точек зрения остается фундаментальным. Кроме всего того, что требуется для регистрации наборов данных по SKAN⁸² или VoID⁸³, присутствует ясно выраженная необходимость в информации не только об онтологии или структуре данных, но и о количестве записей и связей (отношений), частоте обновлений, способе управления обновлениями и об источниках, используемых при слиянии любых данных. Можно сказать, исходя из последнего опыта установления связей данных с внешними источниками, что эти источники (как

⁸² Comprehensive Knowledge Archive Network. [<http://ckan.net/package/new>]

⁸³ Describing Linked Datasets with the VoID Vocabulary [<http://www.w3.org/TR/void/>]

это хорошо знают библиотекари и библиографы) имеют огромный разброс утверждений об одних и тех же вещах: иногда они неточны, иногда – ошибочны, иногда избыточны, а в других случаях порождают недоверие к источникам.

Семантический веб и LOD предлагают увеличить точность навигации, но можно заметить: несмотря на то, что сейчас в LOD доступны лишь несколько источников, можно достаточно просто разными способами найти то, определение чего дано разными способами, с самыми разными атрибутами, но зато одного связывания недостаточно для улучшения точности навигации. Здесь нужно заметить, что уже было сказано о необходимости трансформировать в наборы данных LOD большого объёма библиографических источников высочайшего качества, остающихся на бумаге, неструктурированных и вне зоны общественного доступа. За это отвечают библиотеки и библиотекари, которые не имеют права откладывать это на «потом», - подтверждением тому является продвижение массовых проектов оцифровки – Европейцы и Испанцы. Как известно, немецкий идеализм XIX века учит: *Das Ding im sich, das Ding am sich у das Ding bei sich.*

9. Заключение

Все работы, описанные здесь вращаются вокруг двух проектов, которые на момент завершения последнего обзора - 30 апреля 2011 года, - остаются открытыми. С одной стороны – только что началось внедрение очередной фазы Европейцы – «Дунай». С другой стороны, ожидается, что результат пробного внедрения RDA в ряде американских библиотек станет достоянием общественности, возможно – на ближайшем заседании Американской библиотечной ассоциации. Если добавить, что выводы рабочей группы W3C Группы инкубатора библиотечных связанных данных еще не закончены, то очевидно, что наши результаты - это просто набор действий, параллельных разработке и реализации этой новой парадигмы. Может быть, именно там находится то, за счёт чего наш клад может представлять наибольший интерес, т.к виртуальная библиотека «Полимаг» предприняла шаги, похожие на внедрение RDA и активно участвует в Европеане, что является прецедентом использования LLD. Таким образом, нашу реальную работу, доступную через веб тем, кто хочет её увидеть, можно рассматривать как дополнительный вклад в достижении того, чтобы информация (и не только библиографическая) стала максимально доступной, и, прежде всего, в достижении целей, предложенных Тимом Бернерс-Ли, когда он провозгласил Linked Open Data. Это и есть, в конечном итоге, смысл этой связи.