

ПРЕЗИДЕНТСКАЯ БИБЛИОТЕКА ИМЕНИ Б.Н.ЕЛЬЦИНА

Библиотечные стандарты и Семантическая паутина

Алексей Александрович Воробьев
14 мая 2013г.



СОДЕРЖАНИЕ

1. Миссия и задачи библиотек в современном информационном пространстве.
2. Модель библиографических данных FRBR и стандарт описания ресурсов - RDA.
3. RDF - модель представления данных в Web3.0
4. Схожесть представления моделей FRBR и Semantic Web.
5. Зарубежный и отечественный опыт.
6. Вывод и предложения.

МИССИЯ

Социальная миссия библиотек заключается в сохранении и организации широкого доступа населения к культурному наследию, представленного в виде книг, газет, журналов, карт, аудио-видео ресурсов и электронных ресурсов.

Для выполнения своей цели современная библиотека изменила свои границы, перейдя из реального пространства в виртуальное, в Web среду .

ЗАДАЧИ

1. Согласно программе основных направлений развития библиотечно-информационной сети «ЛИБНЕТ» на 2011-2020 гг. предусматривается поэтапное обеспечение к 2020 г. свободного доступа всех граждан России к цифровой форме любого документа, хранящегося в любой библиотеке страны через Интернет с соблюдением требований авторских прав.
2. К 2020 году планируется создание единой общероссийской библиотечной сети в Web пространстве (к 2015г. Россия должна иметь электронные каталоги в 100% библиотек и 50% документов в цифровой форме, к 2020г. - 75% оцифрованных документов).

ПРОБЛЕМА

Библиотеки размещают свои электронные каталоги и цифровые ресурсы в сети Интернет.

Но:

- как **отдельные базы данных**, которые не связаны между собой и базами других сообществ
- сквозной поиск информации затруднен
- нет единой точки входа
- библиотечные системы не позволяют напрямую интегрировать данные из внешних источников.

ПУТИ РЕШЕНИЯ

Что делать?

Просто выбрать средства для предоставления данных в Web среде, так чтобы они максимально были доступны пользователям, и начать их использовать?

Нет!

**Мало просто разместить данные в Web,
необходимо изначально проектировать их в
контексте Web!**

Модель библиографических данных FRBR

FRBR (Functional Requirements for Bibliographic Records)
Функциональные требования к библиографическим записям.

FRBR – это модель связей различных объектов, представленная как концептуальная структура, позволяющая создавать библиографические записи независимо от любых правил каталогизации и формата представления. Разработчики концепции опирались на важнейшие пользовательские интересы: найти, идентифицировать, выбрать и получить доступ к определенному ресурсу.

Концепция была разработана международным библиотечным сообществом под эгидой Международной федерацией библиотечных ассоциаций и учреждений (IFLA) в 1998 г.

Краткое описание модели FRBR

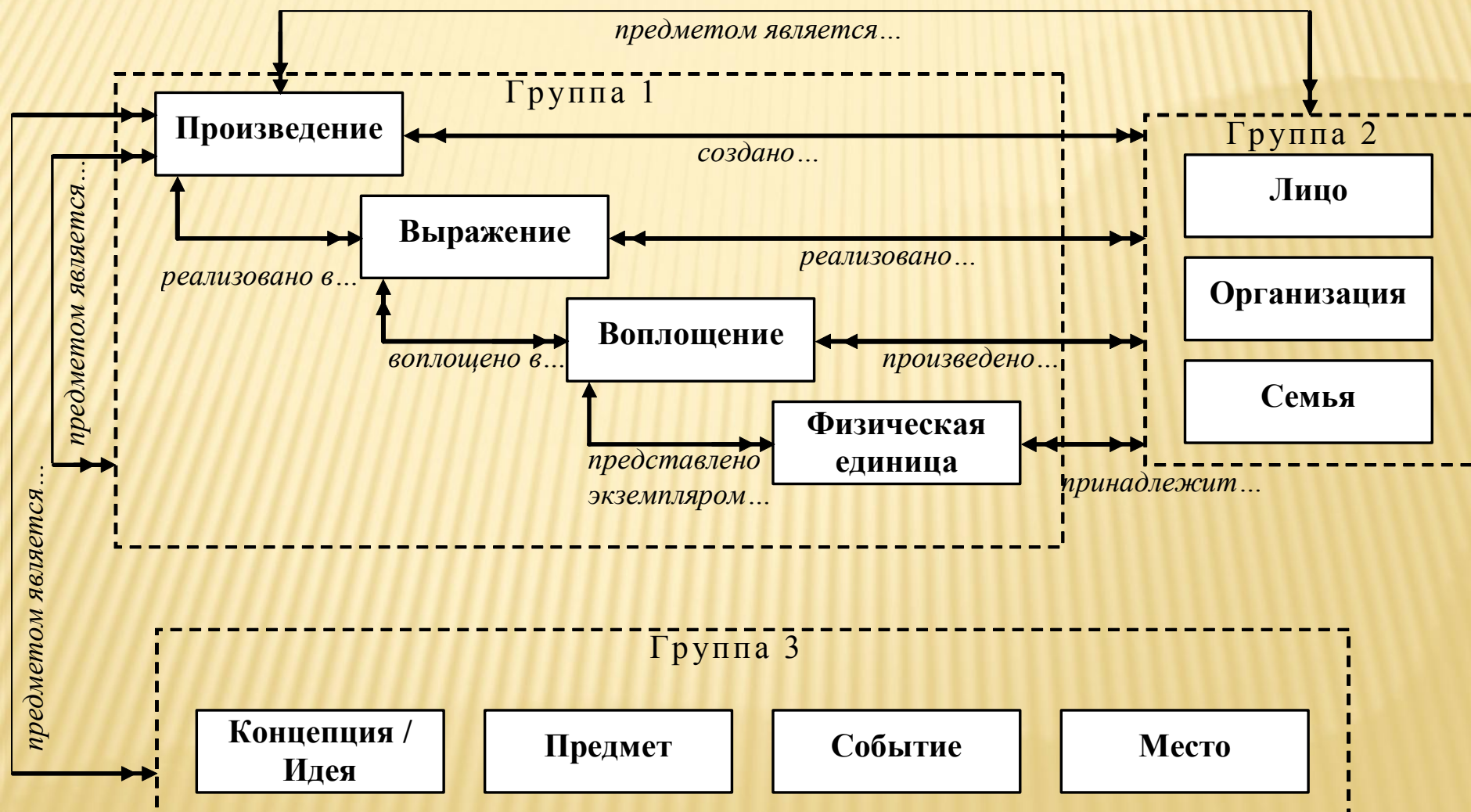
Модель FRBR включает три группы объектов:

Группа 1 - произведение (work), выражение (expression), воплощение (manifestation), физическая единица (item.)

Группа 2 - *лицо (person), организация (corporate body), семья (family)*, которые несут ответственность за интеллектуальное или художественное содержание, производство, распространение, хранение объектов Группы 1 либо владеющие объектами Группы 1.

Группа 3 - *концепция/идея (concept)* – абстрактное понятие или идея, *предмет (object)* – материальная вещь, предмет; *событие (event)* – действие либо период времени; *место (place)* – местонахождение. Группу составляют объекты, являющиеся предметом интеллектуальной или художественной работы. Кроме того, в качестве предмета *произведения* может выступать любой из объектов первой или второй групп. Например: *произведение о другом произведении* или *произведение об отдельной физической единице* – *произведение*, посвященное редкой книге, существующей в единственном экземпляре.

Краткое описание модели FRBR



Краткое описание модели FRBR

Центральными связями для модели FRBR являются связи первой группы. Связи могут быть установлены между различными произведениями и их выражениями, например между произведением А.С. Пушкина «Евгений Онегин», представленного в виде литературного текста и произведением П.И. Чайковского «Евгений Онегин» в виде оперы, между одним произведением и несколькими выражениями, например текст произведения «Евгений Онегин» на русском, английском и немецком языках и т.д.

На основе формирования таких связей мы можем получить в ответ на пользовательский запрос «Евгений Онегин», например, следующий результат:

Поиск: Евгений Онегин

Записей: 49

Автор: Пушкин Александр Сергеевич (1799-1837)

Заглавие: Евгений Онегин

Найдено: (w1) Роман в стихах (текст) (45)
(w2) Опера (1)
(w3) Рецензия/Анализ (3)

СЕМАНТИЧЕСКАЯ ПАУТИНА

Семантическая паутина (СП) - это направление развития Всемирной паутины, целью которого является принципиально новый подход в описании ресурсов, смысл (семантика) которых должен быть «понятен» не только людям, но и компьютерным программам.

Идеолог: Тим Бернерс-Ли, один из основоположников всемирной паутины (WWW) и директор WWW-консорциума (W3C).

Консорциум Всемирной паутины (W3C)

- **Консорциум Всемирной паутины (World Wide Web Consortium, W3C)** — организация, разрабатывающая и внедряющая технологические стандарты для Всемирной паутины в сети Интернет.
- Консорциум был создан в 1994 году как консультативный орган для лидеров компьютерной индустрии.
- Миссия W3C: «Полностью раскрыть потенциал Всемирной паутины, путём создания протоколов и принципов, гарантирующих долгосрочное развитие Сети».
- Членами консорциума являются более 350 участников из 28 стран.

W3C И БИБЛИОТЕКИ

- 2010 год организована рабочая группа по выработке рекомендаций представления библиотечных данных в Web среде.
- 2011 год рабочая группа подготовила 3 отчета по библиотечным связанным данным.
- 16 февраля 2012 года открыто представительство в России.

ОСНОВНАЯ ИДЕЯ СП

- вся информация должна быть описана в соответствии с моделью RDF (Resource Description Framework) триплетом

Субъект-Отношение-Объект

- **URI** (англ. *Uniform Resource Identifier*) — унифицированный (единообразный) идентификатор ресурса.
- данные должны быть **открытые**, т.е. доступные всем пользователям WWW и **связанные**, т.е. иметь многосторонние связи с другими данными

Важнейший элемент Semantic Web - онтологии

Все элементы семантической паутины должны быть зарегистрированы и описаны с помощью онтологий для обеспечения технической и смысловой интероперабельности (способности к взаимодействию) различных систем.

Регистрация словарей осуществляется на специализированных сайтах, например:

- <http://dublincore.org/>
- <http://metadatatagistry.org/>

The image shows two web browser windows side-by-side. The left window displays the Dublin Core Metadata Initiative website, which includes a search bar, navigation tabs (Home, Metadata Basics, DCMI Specifications, Community and Events), and a sidebar with links to Twitter and featured glossary entries. The right window displays the Open Metadata Registry website, featuring a 'Welcome to The Registry!' message, a 'How to begin' section, and a 'Registry News' section with a table of recent updates. The table lists updates to the National Library of Scotland Digital Object Database elements, including changes to keywords, labels, and URIs.

Registry News ...from the Registry Blog	Latest Activity
Whoops Posted by: Jon at 23:32 on Wednesday, September 28, 2011 GMT Until lately we've been pretty happy with our ISP, Dreamhost. But a few months ago, after several during-the-presentation meltdowns of the Registry we determined that we needed to move to a higher-performing, more reliable server. We could have done the easy thing and moved to a Virtual Private Server at Dreamhost. Instead, we setup an [...]	National Library of Scotland Digital Object Database elements :: has keyword :: label :: deleted On Sat Sep 29 2012 at 20:09:26 EDT, Gill Hamilton, working on the Element: has keyword , deleted the label (en) Property: http://nlsdata.info/dod/elements/keywordOf
SPARQL queries Posted by: Jon at 14:55 on Thursday, April 14, 2011 GMT We're using Benjamin Nowack's excellent ARC libraries for a tiny bit of our RDF management. It may surprise you to know that we don't use a triple store as our primary data store, but we do too many things with the data that we think are cumbersome at best when managed exclusively in a triple store (a subject [...])	National Library of Scotland Digital Object Database elements :: has place created :: url :: updated On Sat Sep 29 2012 at 18:44:55 EDT, Gordon Dunsire, working on the Element: has place created , updated the url (en) Property To: http://nlsdata.info/dod/elements/PlaceType-Placecreated From: http://nlsdata.info/dod/elements/PlaceType-Placecreated
Check Your Bookmarks! Posted by: Diane Hillmann at 19:20 on Tuesday, October 05, 2010 GMT As of today, the link to Step-by-Step instruction on the Registry front page has been	

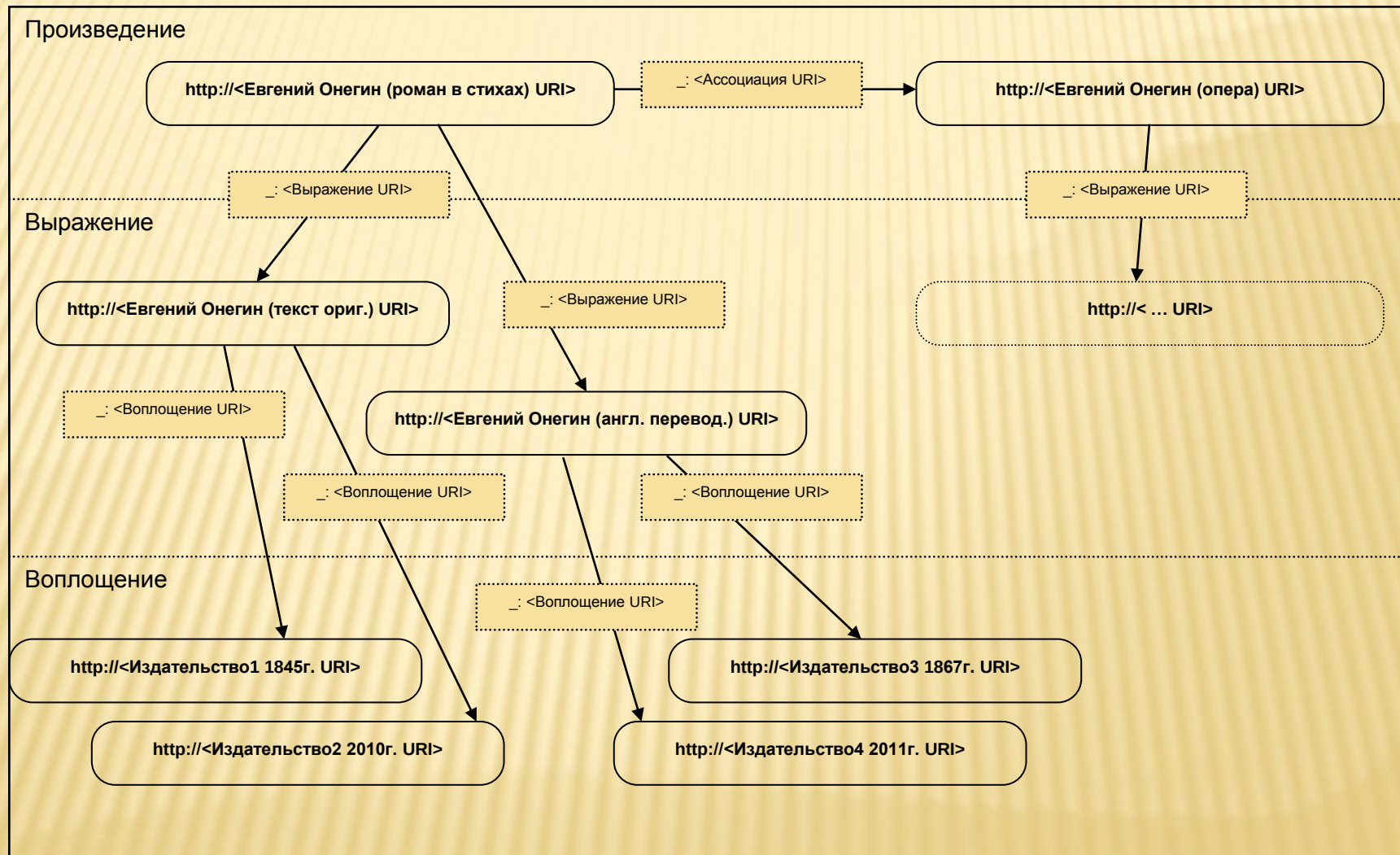
Пример RDF триплета

http://<Евгений Онегин (роман в стихах) URI>

http://purl.org/dc/elements/1.1/creator

http://<Пушкин Александр Сергеевич (1799-1837) URI>

Модель FRBR в концепции семантического Web



ЗАРУБЕЖНЫЙ ОПЫТ

Теория или практика?

-Крупные компании, такие как Microsoft, IBM, Adobe, Sun Microsystems, Google и другие активно используют технологию Семантического Web в своих продуктах для решения задач управления данными.

-Библиотечное сообщество также находится на переходном этапе от теории к практике. На основе концептуальной модели FRBR американскими библиотечными специалистами разработан новый стандарт описания ресурсов - RDA (Resource Description and Access - Описание ресурсов и доступ к ним).

Зарубежный опыт Стандарт описания RDA

RDA содержит набор правил на основе модели FRBR.

В 2008 – 2011 гг. было проведено тестирование нового стандарта описания в американских библиотеках. По результатам тестирования было принято решение о внедрении нового стандарта описания в американских библиотеках, которое начнется в марте 2013 года.

В качестве новой структуры представления информации о библиотечных ресурсах предполагается использовать структуру описания *RDF*, являющуюся базовой структурой описания данных в семантическом Web.

Уже сейчас ряд библиотек мира работают над переводом своих электронных каталогов, представленных в традиционных MARC форматах в структуры *RDF*.

Наиболее масштабный проект – это представление национальной британской библиографии в *RDF* структуре. Было переведено около 3 миллионов библиографических записей.

ЗАРУБЕЖНЫЙ ОПЫТ

- ***Национальная библиотека Великобритании***

Британская Национальная библиография (порядка 3 млн. записей)

- ***Библиотека Конгресса***

полностью перевела авторитетные данные

- ***Национальная библиотека Германии***

библиографические и авторитетные данные

и другие..

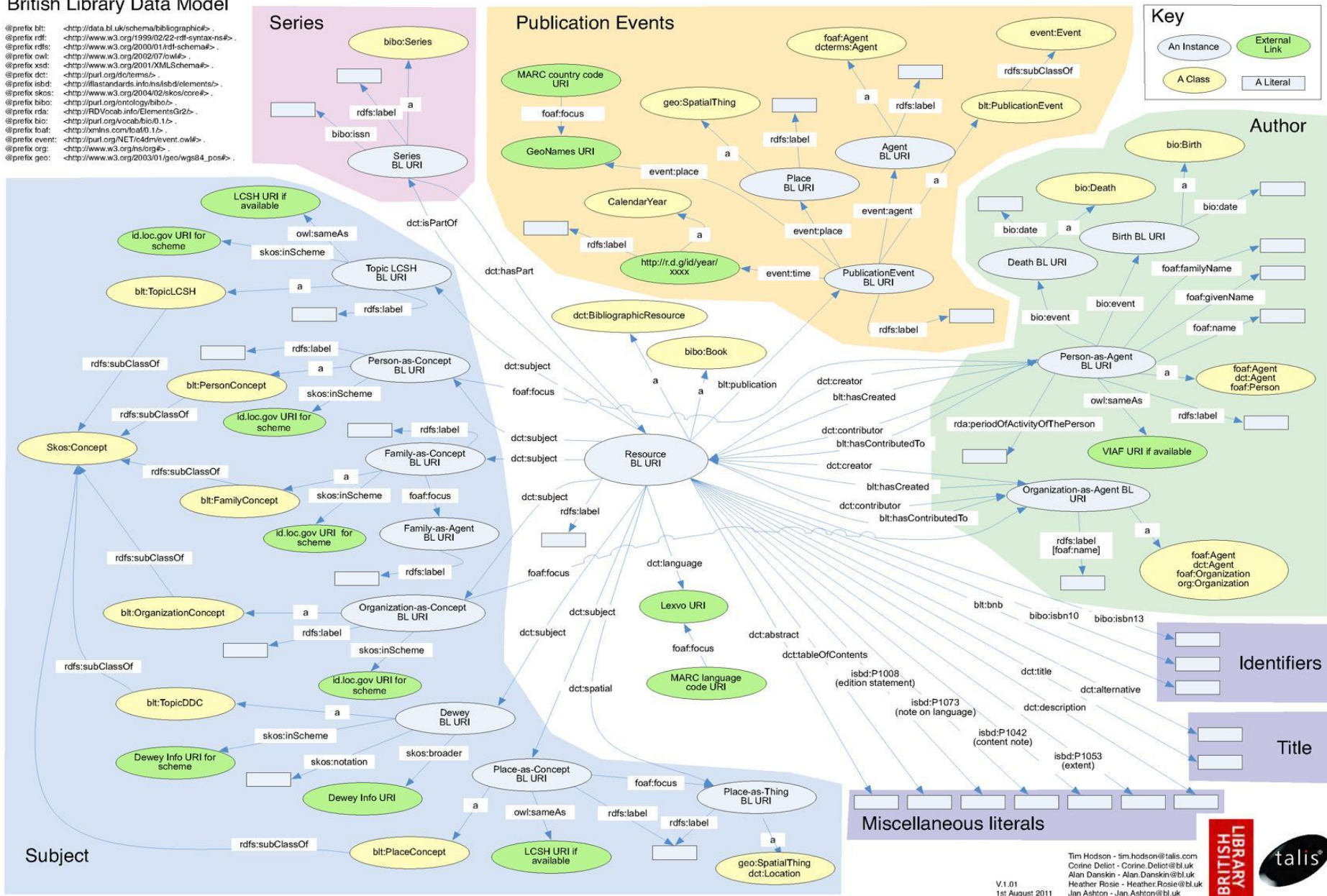
Модель данных для книг Британской национальной библиографии

British Library Data Model

```

@prefix bib: <http://data.bib.org/urn:isbn:085631984>;
@prefix rdf: <http://www.w3.org/1999/02/22/rdf-syntax-ns#>;
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>;
@prefix owl: <http://www.w3.org/2002/07/owl#>;
@prefix xsd: <http://www.w3.org/2001/XMLSchema#>;
@prefix dct: <http://purl.org/dc/terms/>;
@prefix isbd: <http://liststandards.info/isbd/elements/>;
@prefix skos: <http://www.w3.org/2004/02/skos/core#>;
@prefix bibo: <http://purl.org/ontology/bibo/>;
@prefix rda: <http://rdvocab.info/ElementsGr2/>;
@prefix bibc: <http://purl.org/kocab/cite/0.1/>;
@prefix bibl: <http://purl.org/ontology/bibl/>;
@prefix event: <http://purl.net/Net44/event/event.owl#>;
@prefix org: <http://www.w3.org/2003/org/>;
@prefix geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>;

```



Примеры словарей для библиографического описания материалов Британской библиотеки

Библиографические данные описаны с использованием следующих словарей:

- ❖ Bibliographic Ontology
- ❖ Vocabulary for Biographical Information
- ❖ British Library Terms
- ❖ RDA vocabulary
- ❖ Dublin Core
- ❖ Event Ontology
- ❖ Friend of a Friend
- ❖ ISBD
- ❖ Organization Ontology
- ❖ OWL, SKOS, RDF Schema, WGS84 Geo Positioning и других.

Для описания российских библиотечных материалов технология Semantic Web не используется, словарей на русском языке нет.

Библиотека Конгресса США

Библиотека Конгресса, начиная с 2006 года, приступила к представлению своих авторитетных данных в RDF SKOS (Simple Knowledge Organization System Система организации знаний) .

В настоящее время служба поддержки авторитетных файлов Библиотеки Конгресса называется “LC Linked Data Service Authorities and Vocabularies” (Служба связанных данных Библиотеки Конгресса авторитетные файлы и словари).

Таким образом можно сказать, что библиотека также готова представлять свои данные в семантическом Web.

The screenshot shows the Library of Congress website. At the top, there is a navigation bar with the Library of Congress logo and links for "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS". A search bar is also present. Below this, a breadcrumb trail reads "The Library of Congress > Linked Data Service". The main content area features the title "LC Linked Data Service Authorities and Vocabularies" in a large, bold font, which is circled in red. To the left of this title is a sidebar with a list of links: "Linked Data Service", "About", "Dataset Descriptions", "Search", "Download", "Technical Center", and "Contact Us". Below the title is a decorative image of a classical architectural frieze. At the bottom of the page, there is a search section with a search bar, a dropdown menu for selecting the search scope (currently set to "All"), and buttons for "Search" and "Reset". A note on the right side of the search section states: "**Please Note: LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the scheme. This is temporary while the impact of adding LCC to the current system is better understood."

Список предметных рубрик Библиотеки Конгресса

Details Visualization Suggest Terminology

 **Azov, Sea of (Ukraine and Russia)**

URI(s)

- > <http://id.loc.gov/authorities/subjects/sh85010667>
- > <info:lc/authorities/sh85010667>
- > <http://id.loc.gov/authorities/sh85010667#concept>

Instance Of

- > [MADS/RDF Geographic](#)
- > [MADS/RDF Authority](#)
- > [SKOS Concept](#)

Scheme Membership(s)

- > [Library of Congress Subject Headings](#)

Collection Membership(s)

- > [LCSH Collection - Authorized Headings](#)
- > [LCSH Collection - General Collection](#)

Variants

- >  Azof, Sea of (Ukraine and Russia)
- >  Azov, Sea of (Ukraine and R.S.F.S.R.)
- >  Azovskoe more (Ukraine and Russia)
- >  Azovskoye more (Ukraine and Russia)
- >  Maeotis Palus (Ukraine and Russia)
- >  Palus Maeotis (Ukraine and Russia)
- >  Sea of Azof (Ukraine and Russia)
- >  Sea of Azov (Ukraine and Russia)

Broader Terms

- >  [Seas--Russia](#)
- >  [Seas--Ukraine](#)

Earlier Established Forms

- > Azov, Sea of (Ukraine and R.S.F.S.R.)

Sources

- > found: Mineralogiia Priazov'ia, 1981.
- > found: BGN, Nov. 13, 1981 (Azovskoye More [Russian]; Sea of Azov [Conventional]; sea: 46°00'N, 36°00'E; variants: Maeotis Palus; Sea of Azof; Azovskoe More)
- > found: Old catalog heading (Azov, Sea of)
- > found: Web. geog. (Azov, Sea of or Sea of Azof or Russ. Azovskoye More, or anc. Palus Maeotis)
- > found: Lippincott (Azov, Sea of, anc. Palus Maeotis, shallow Narm of Black Sea)
- > found: BGN sea; 46°00'00"N 036°00'00"E

Change Notes

- > 2003-10-01: new
- > 2011-06-06: revised

Alternate Formats

- > [RDF/XML \(MADS and SKOS\)](#)
- > [N-Triples \(MADS and SKOS\)](#)
- > [JSON \(MADS/RDF and SKOS/RDF\)](#)
- > [MADS - RDF/XML](#)
- > [MADS - N-Triples](#)
- > [MADS/RDF - JSON](#)
- > [SKOS - RDF/XML](#)
- > [SKOS - N-Triples](#)

Пример предметной рубрики
«Азовское море».

Информация может быть доступна в
любом из форматов семантического Web.

Пример авторитетной записи предметной рубрики «Азовское море» в RDF (формат N-Triples) Библиотеки Конгресса США

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#prefLabel> "Azov, Sea of (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2008/05/skos-xl#altLabel> _:bnode7authoritiessubjectssh85010667 .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2008/05/skos-xl#altLabel> _:bnode12authoritiessubjectssh85010667 .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Azof, Sea of (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Azov, Sea of (Ukraine and R.S.F.S.R.)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Azovskoe more (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Azovskoye more (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Maeotis Palus (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Palus Maeotis (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Sea of Azof (Ukraine and Russia)"@en .

<http://id.loc.gov/authorities/subjects/sh85010667> <http://www.w3.org/2004/02/skos/core#altLabel> "Sea of Azov (Ukraine and Russia)"@en .

...

РОССИЙСКИЙ ОПЫТ

Опыта нет!

Библиотечное сообщество еще практически не включилось в этот процесс.

Текущее состояние российских библиотек

К настоящему времени российские библиотеки обладают значительными массивами метаданных и контролируемых словарей, переведенных в MARC форматы.

Объем Сводного каталога библиотек России составляет около 7 мил. библиографических записей на ресурсы, хранящихся в библиотеках России и около 2 мил. авторитетных записей.

Однако, эта информация не доступна для широкого круга пользователей и поиск ее затруднен.

Выводы

Для обеспечения свободного доступа к информации российскому библиотечному сообществу требуется воспользоваться опытом зарубежных коллег по представлению библиотечных данных в Web.

Социальная значимость и масштабность задачи по переводу библиографических данных страны, используя новую модель RDF в Web среде, требует государственную поддержку.

Практические шаги и планы

1. В Президентской библиотеке создана рабочая группа, включающей специалистов в области библиотечного дела и информационных технологий для разработки и внедрения стандартов Семантической паутины для российских библиотек.
2. Идет подготовка к практической реализации представления тестового массива данных (российские государственные деятели) по стандартам Семантической паутины.

ЦЕЛИ И ЗАДАЧИ РАБОЧЕЙ ГРУППЫ

Основная цель – обеспечение представления российских библиотечных данных средствами Семантической паутины (СП).

Задачи:

- ❖ Создание технологических и программных решений по преобразованию (конвертации) библиотечных данных из формата RUSMARC в форматы, совместимые с RDF.
- ❖ Технологическое обеспечение публикации библиографических данных в формате открытых данных, в сети Интернет.
- ❖ Создание онтологий и словарей контролируемой лексики.
- ❖ Модернизация технологий поиска и других пользовательских сервисов.

ОЖИДАЕМЫЕ РЕЗУЛЬТАТЫ

- ✓ Разработка концепции представления российских библиотечных данных в СП.
- ✓ Разработка таблицы конвертирования данных из библиотечных форматов в форматы СП (библиографические и авторитетные данные).
- ✓ Создание конвертеров данных формата RUSMARC в форматы, совместимые с СП.
- ✓ Тестирование конвертеров.
- ✓ Экспериментальная публикация сконвертированных данных в сети Интернет как открытых данных.

СПАСИБО ЗА ВНИМАНИЕ!

www.prlib.ru
vorobyov@prlib.ru