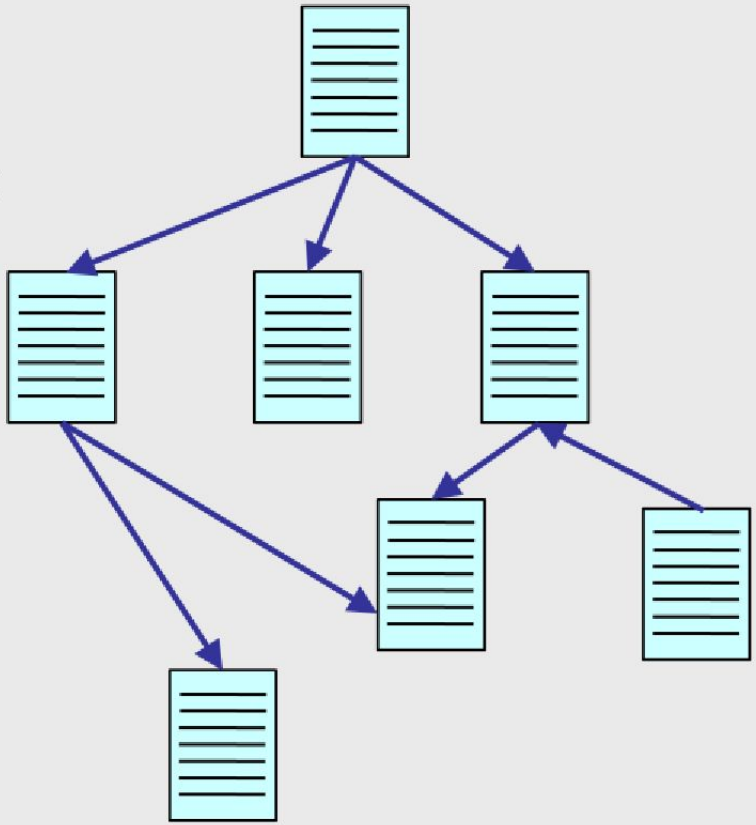
Перспективы развития каталогизации: от MARC-форматов к связанным данным

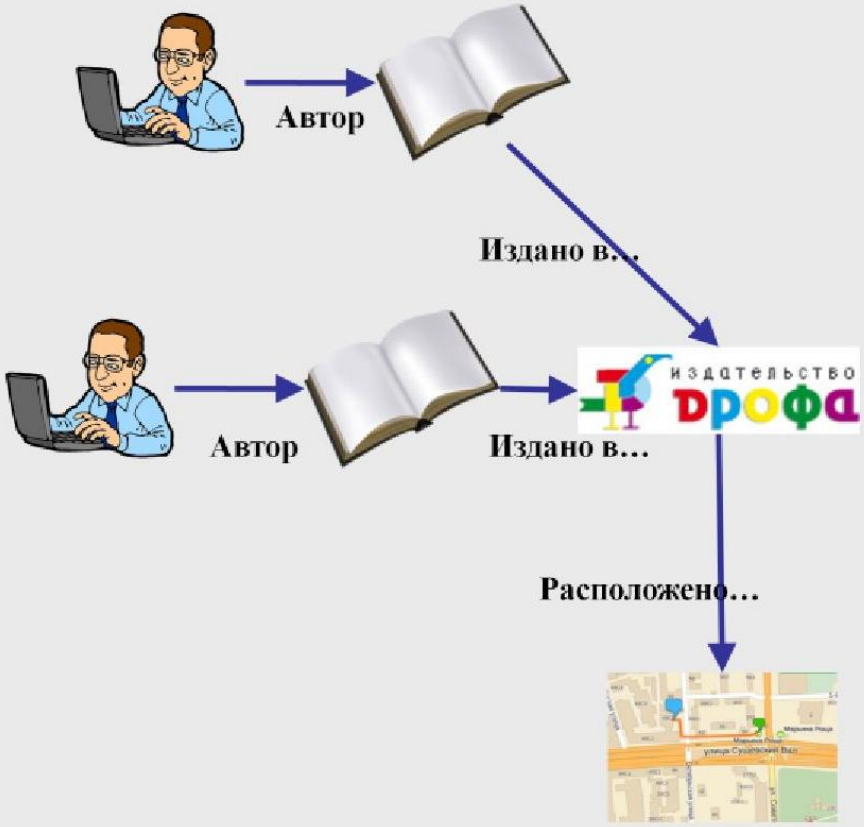
МГОУНБ

Web 2.0

- Сеть документов (сайтов, web-страниц), связанных посредством гиперссылок
- Связи устанавливаются между текстами документов (веб-страницами, файлами)
- Для установки ссылок используются универсальные идентификаторы ресурсов (URI).
- Связи односторонние – в документе нет никакой информации о том, что на него ссылается другой документ
- Связи неинформативные – говорят только о наличии связи, но не о ее характере.



Semantic Web (Web 3.0)



- Надстройка над Всемирной паутиной; сеть данных, обеспечивающая обработку и интерпретацию данных компьютерами
- Связи устанавливаются между объектами (данными)
- Связи информативные – говорят не только о наличии связи, но не о ее характере.
- Для установки ссылок используются универсальные идентификаторы ресурсов (URI).
- URI существуют не только у страниц, но и у объектов или абстрактных понятий.
- URI глобально уникальны, поэтому могут использоваться для установления ссылок в разных местах в Semantic Web.

ОСНОВНАЯ ИДЕЯ СП

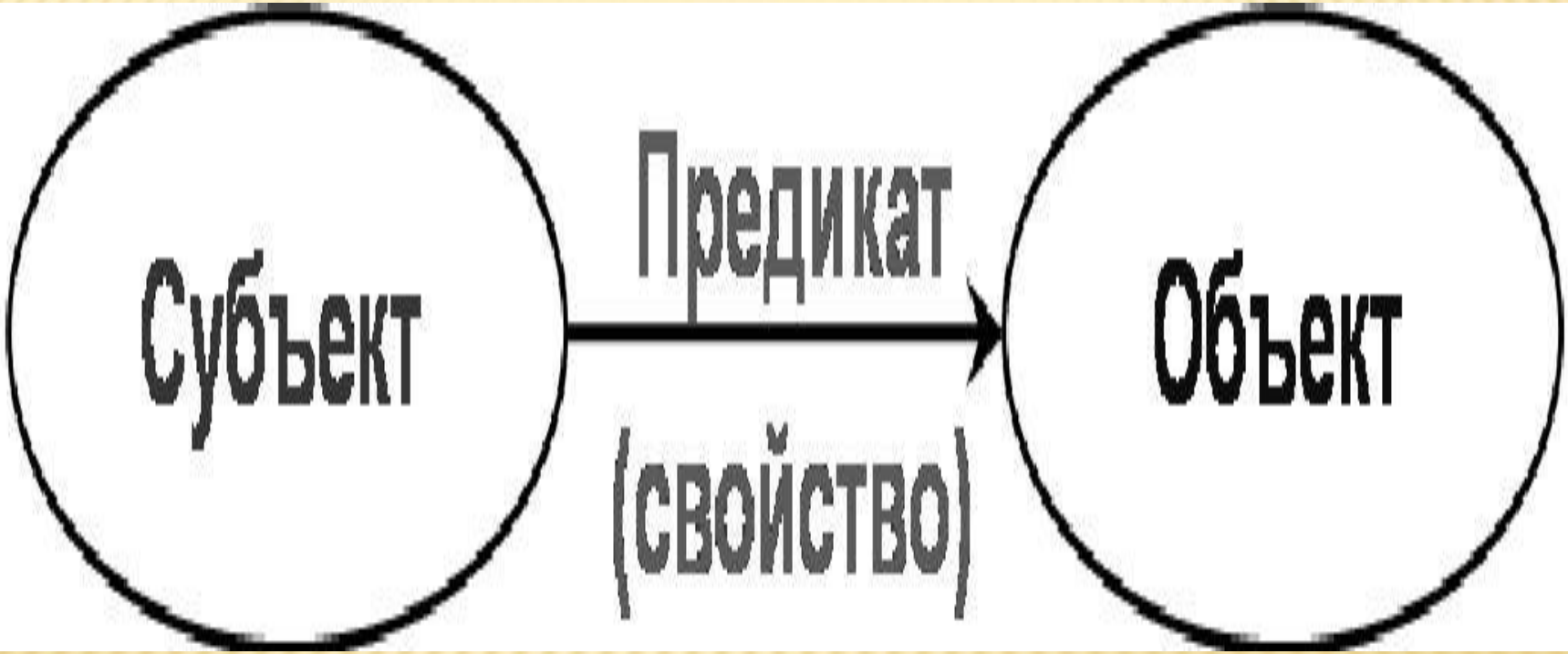
- вся информация должна быть описана в соответствии с моделью RDF (Resource Description Framework) триплетом

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

ТРИПЛЕТ RDF



Множество RDF-утверждений образует ориентированный граф, в котором вершинами являются субъекты и объекты, а рёбра помечены предикатами, например:

Файл Правка Вид Журнал Закладки Инструменты Справка

Национальный информационно- | linked open data — Яндекс: на | The Linked Open Data Cloud

https://lod-cloud.net

Поиск

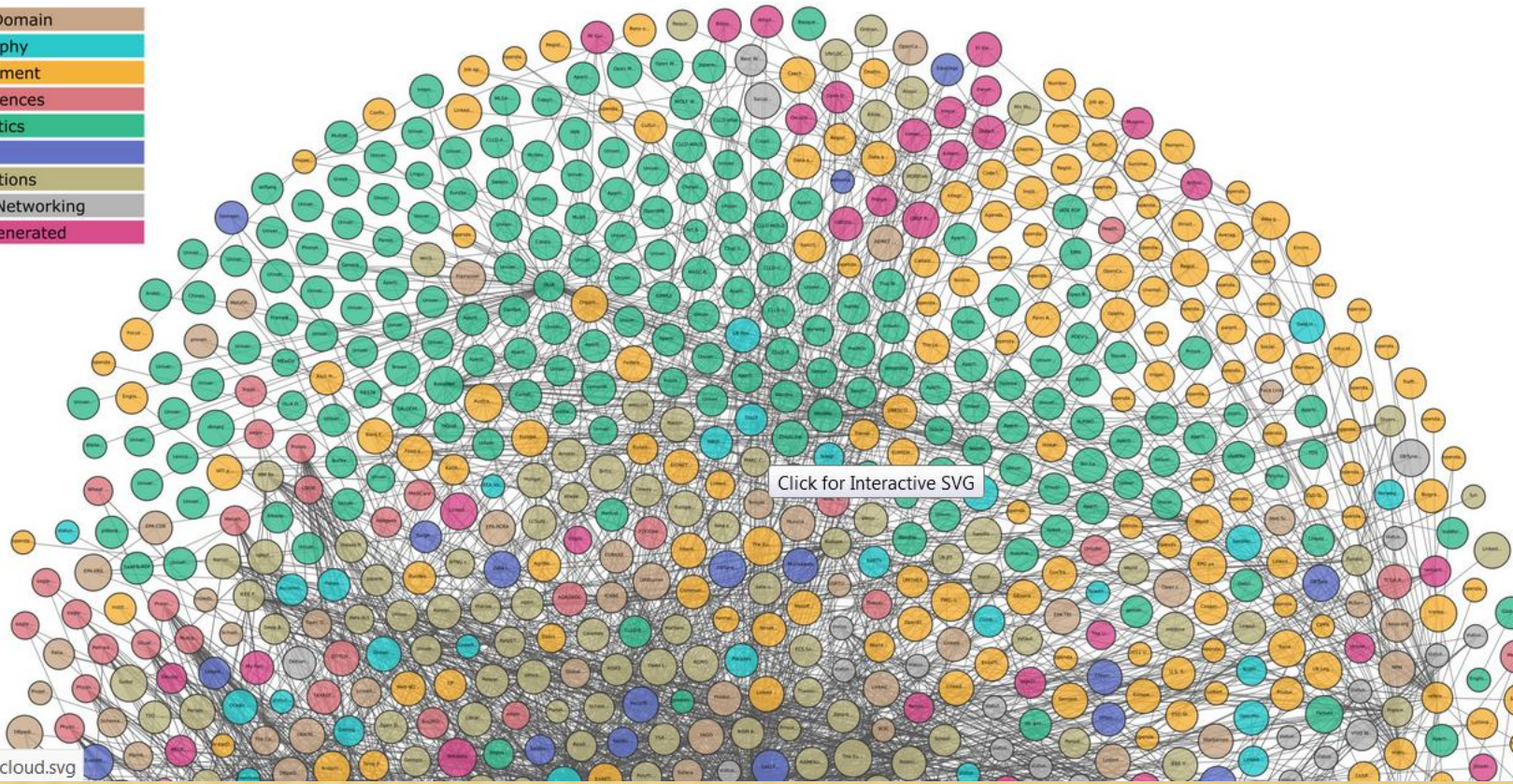
Часто посещаемые | Catálogo de la Bibliot... | Каталог Библиотеки ... | Бесплатный перевод... | Google | «Одежда по размеру... | Электронный катало... | Электронный этало

Browse Submit a dataset Diagram Subclouds IPFS cloud Ab

The Linked Open Data Cloud

Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated



https://lod-cloud.net/clouds/lod-cloud.svg

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

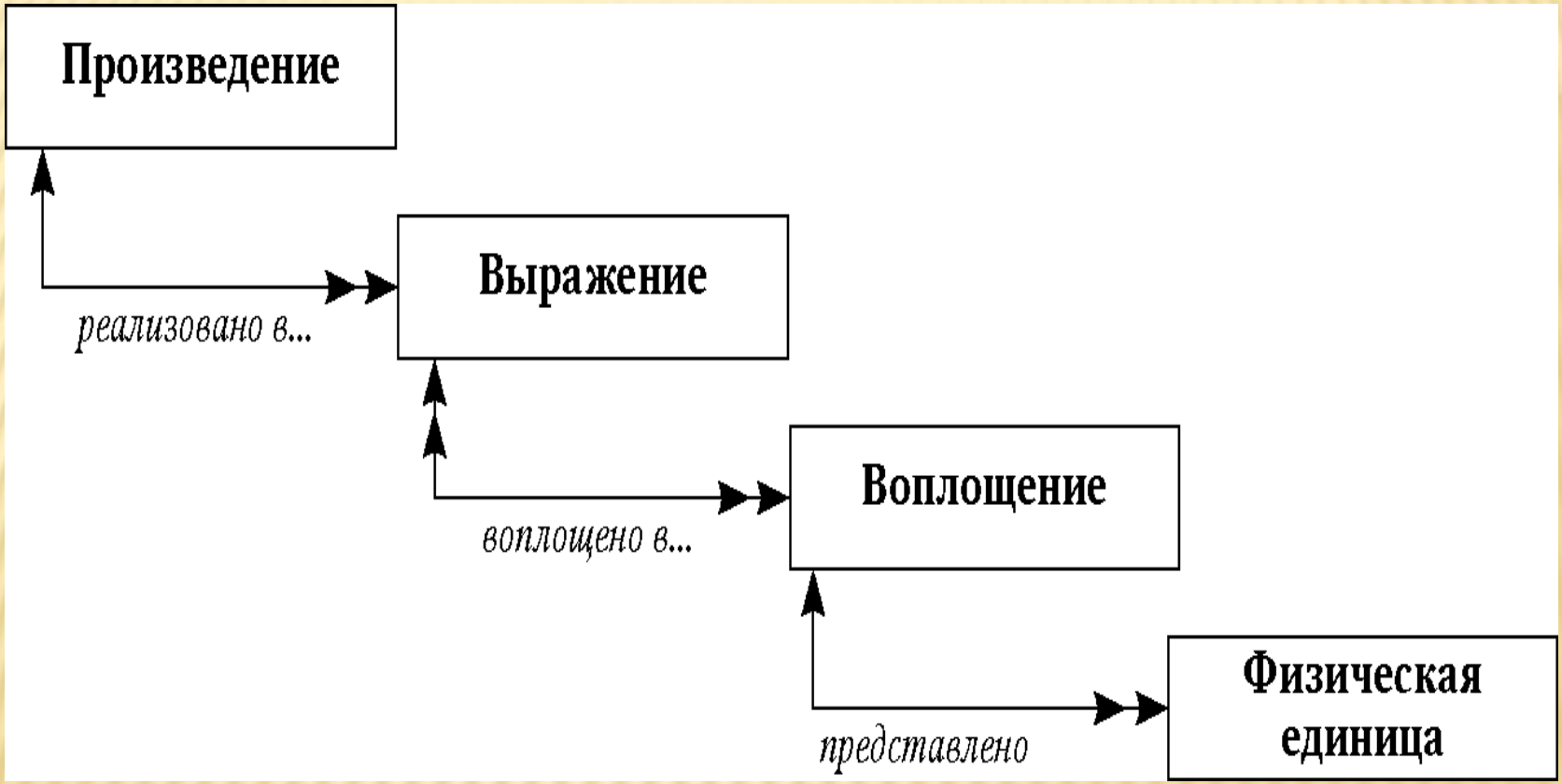
Модель FRBR включает три группы объектов:

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

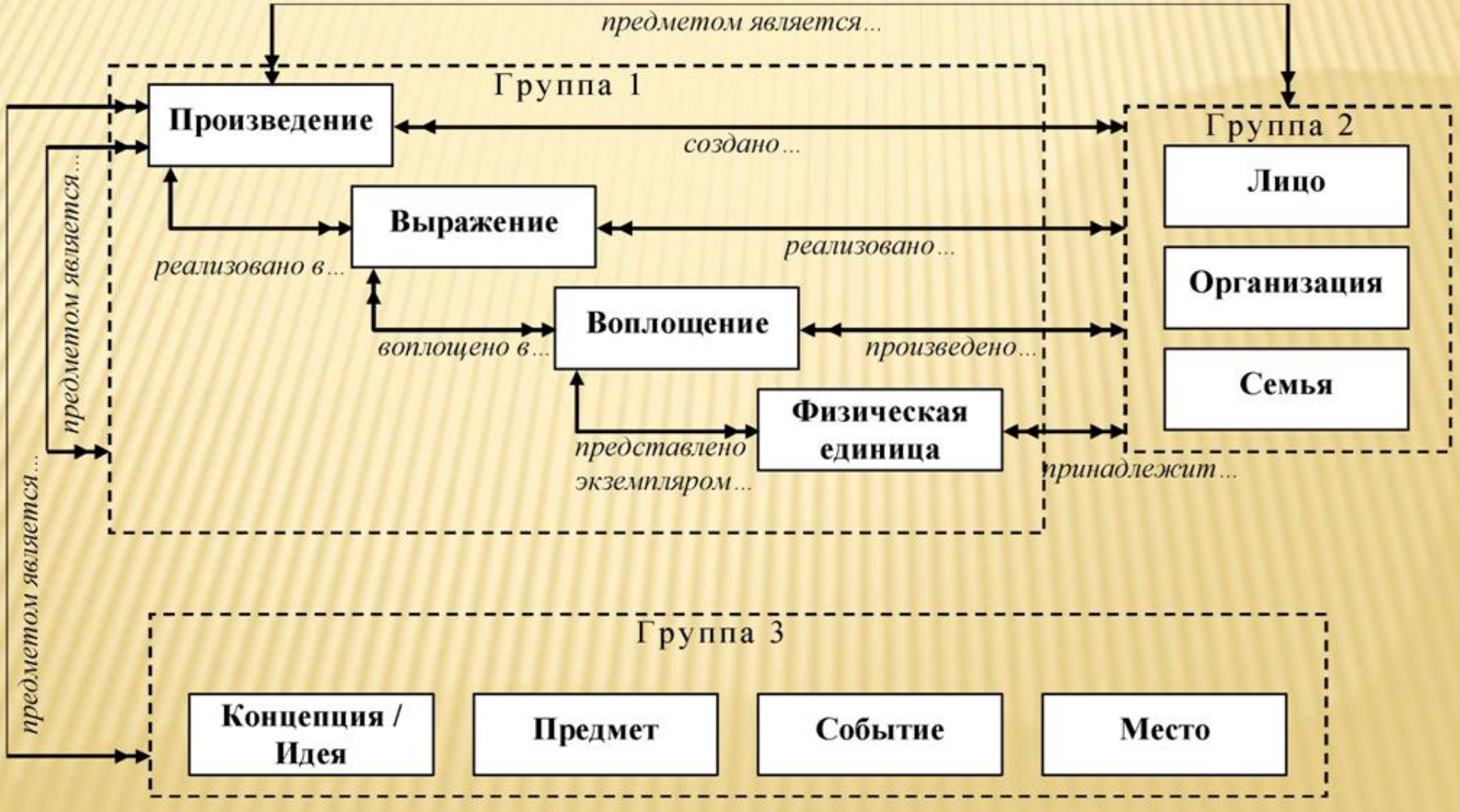
Группа 2 - *лицо (person), организация (corporate body), семья (family)*, которые несут ответственность за интеллектуальное или художественное содержание, производство, распространение, хранение объектов Группы 1 либо владеющие объектами Группы 1.

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.

ТАКИМ ОБРАЗОМ, ПРОИЗВЕДЕНИЕ РЕАЛИЗУЕТСЯ ПОСРЕДСТВОМ ВЫРАЖЕ-НИЯ, ВЫРАЖЕНИЕ - ПОСРЕДСТВОМ ВОПЛОЩЕНИЯ, ПРОЯВЛЕНИЕМ ОТДЕЛЬ-НОГО ВОПЛОЩЕНИЯ ЯВЛЯЕТСЯ КОНКРЕТНЫЙ РЕСУРС



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 нескольким 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)

Реализация модели FRBR

Определение атрибутов объектов, установление библиографически значимых связей между объектами различного уровня и отражение этих связей в каталоге позволяет помочь пользователю сформулировать запрос; наглядно представить взаимосвязи библиографических объектов в структурированном виде; осуществить навигацию в информационном пространстве вообще и в электронном каталоге в частности. В случае реализации модели электронный каталог станет не просто совокупностью библиографических записей, раскрывающей состав и содержание фонда библиотеки, и во многом повторяющей традиционный карточный каталог, а настоящей сетью взаимосвязанных библиографических данных, в которой пользователь может проводить поиск, переходя от одного узла-объекта к другому. Такой способ представления информации, хранящейся в библиотечных каталогах, сделает каталог наиболее полезным и актуальным для пользователя.

Один из возможных вариантов представления результатов поиска при условии реализации модели FRBR. При поиске по заглавию «Тихий Дон» пользователь получает список *произведений*

Поиск: **ТИХИИ ДОН** Записей: **40**

Автор: Шолохов Михаил Александрович (1905-1984)

Заглавие: **Тихий Дон**

Найдено: (wl) Роман (текст) (22)

(w2) Фильмы (3)

(w3) Спектакли (1)

(w4) Опера (1)

(w5) Рецензии / Анализ (3)

ВЫБРАВ ***ПРОИЗВЕДЕНИЕ (W),*** ПОЛЬЗОВАТЕЛЬ ПОЛУЧАЕТ ИЕРАРХИЧЕСКИЙ СПИСОК ***ВЫРАЖЕНИЙ (E)*** И ***ВОПЛОЩЕНИЙ (T)*** КОНКРЕТНОГО ***ПРОИЗВЕДЕНИЯ.*** В ДАННОМ ПРИМЕРЕ ВЫБРАН РОМАН «ТИХИЙ ДОН», ДЛЯ КОТОРОГО В КАТАЛОГЕ ИМЕЮТСЯ СЛЕДУЮЩИЕ ***ВЫРАЖЕНИЯ***: ОРИГИНАЛЬНЫЙ ТЕКСТ НА РУССКОМ ЯЗЫКЕ, ПЕРЕВОДЫ НА АНГЛИЙСКИЙ И КИРГИЗСКИЙ ЯЗЫКИ, АУДИОЗАПИСИ.

(W) ТИХИЙ ДОН [текст]

(e1) Оригинальный текст (рус)
- (m)1. Тихий Дон. - М. : ЭКСМО-Пресс, 2000.
- (m)2. Тихий Дон. - СПб. - Кристалл : Респекс, 1998.
- (m)3. [html]
- (m)4. [rtf]

(e2) Перевод на английский язык
- (m)1. And Quiet Flows Don / transl. by St.Garry. - New York : Vintage Books, 1989.
- (m)2. And Quiet Flows Don / transl. by St.Garry. - New York : A.A. Knopf, 1934.

(e3) Перевод на киргизский язык
- (m)1. Тихий Дон. - Фрунзе : Кыргызстан, 1977-1980.

(e4) Аудиозаписи
- (m)1. [mp3] Тихий Дон / читает Е.Терновский.

ВЫБРАВ НУЖНОЕ ***ВОПЛОЩЕНИЕ*** (В ДАННОМ СЛУЧАЕ - ИЗДАНИЕ 2000 Г., ВЫПУЩЕННОЕ ИЗДАТЕЛЬСТВОМ ЭКСМО-ПРЕСС), ПОЛЬЗОВАТЕЛЬ ПОЛУЧАЕТ РАЗВЕРНУТУЮ ИНФОРМАЦИЮ О ***ВОПЛОЩЕНИИ*** С УКАЗАНИЕМ МЕСТОНАХОЖДЕНИЯ И ШИФРА ***ФИЗИЧЕСКОЙ ЕДИНИЦЫ***

Автор:	Шолохов, Михаил Александрович (1905-)
Основное заглавие:	Тихий Дон
Свед.,отн.к загл.:	роман
Свед.,отн.к загл.:	в 2 кн.
Ответственность:	Михаил Шолохов
Место издания:	Москва
Издательство:	ЭКСМО-Пресс
Дата издания:	2000
Осн.загл.серии:	Русская классика. XX век
Формат:	21
Индекс ББК:	Ш6(2=Р)7-440.4
Язык текста:	rus
Первич.каталогизации:	NLR
Организация:	NLR
Шифр:	2004-3/1866

АНАЛОГИЧНАЯ ИНФОРМАЦИЯ МОЖЕТ БЫТЬ ПОЛУЧЕНА ДЛЯ ДРУГОГО ***ПРОИЗВЕДЕНИЯ*** - ФИЛЬМА «ТИХИЙ ДОН»:

(W) ТИХИЙ ДОН [фильм]

по роману М.А.Шолохова

(e1) Тихий Дон / реж. И. Пырьев. 1930

(e2) Тихий Дон / реж. С. Герасимов. 1958

(m)1. [VHS] Тихий Дон. Подарочное издание (3 кассеты). Спайр

(m)2. [VHS] Тихий Дон. Крупный план / Лизард

(m)3. [VHS] Тихий Дон. Дом Видео

(m)4. [VHS] Тихий Дон. Мастер Тэйп

(m)5. [DVD] Тихий Дон. Коллекционное издание (4 диска). RUSCICO

(e3) Тихий Дон / реж. С. Бондарчук. 2006

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 использоваться 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..

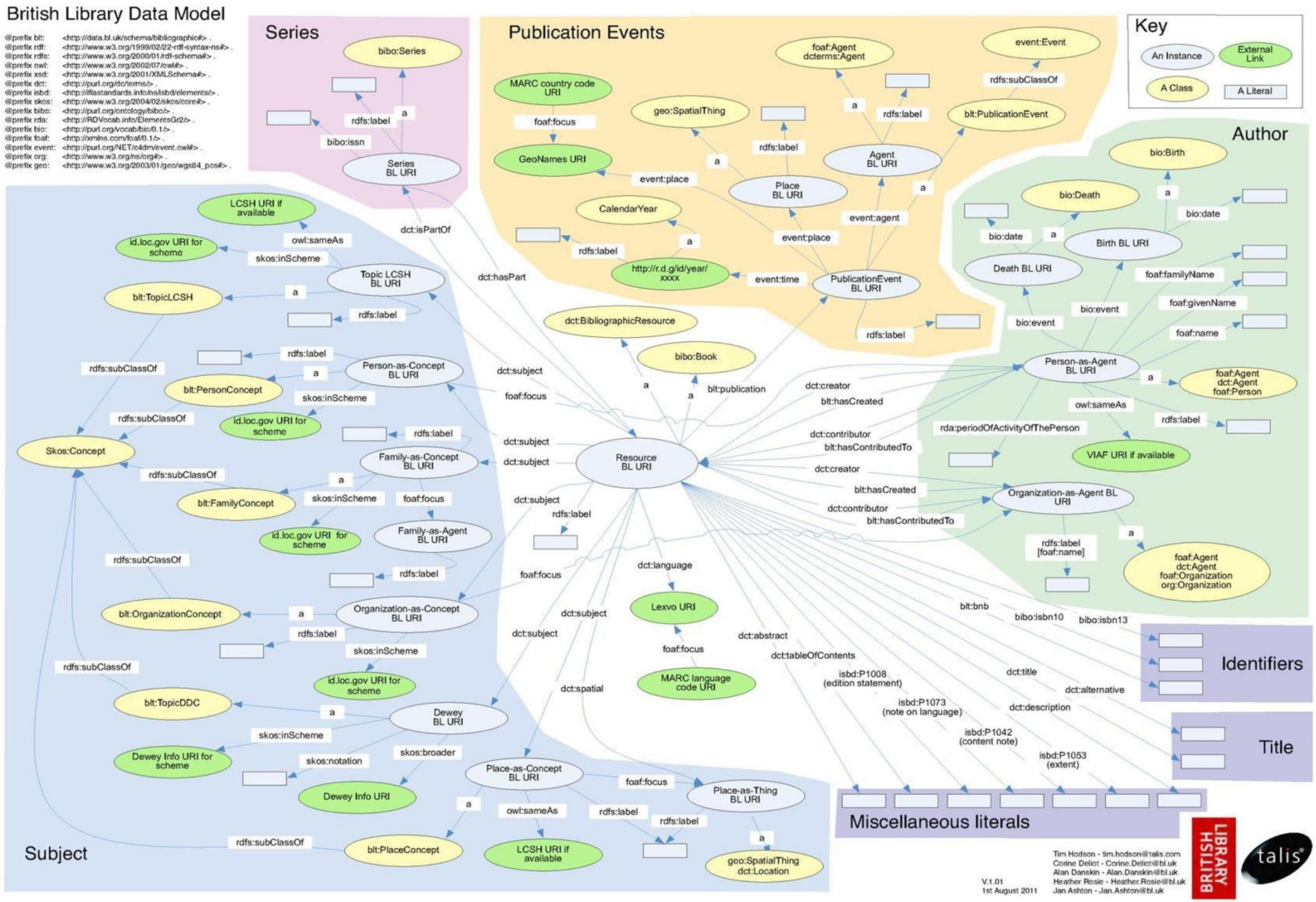
Примеры словарей для библиографического описания материалов Британской библиотеки

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.

Модель данных для книг Британской национальной библиографии

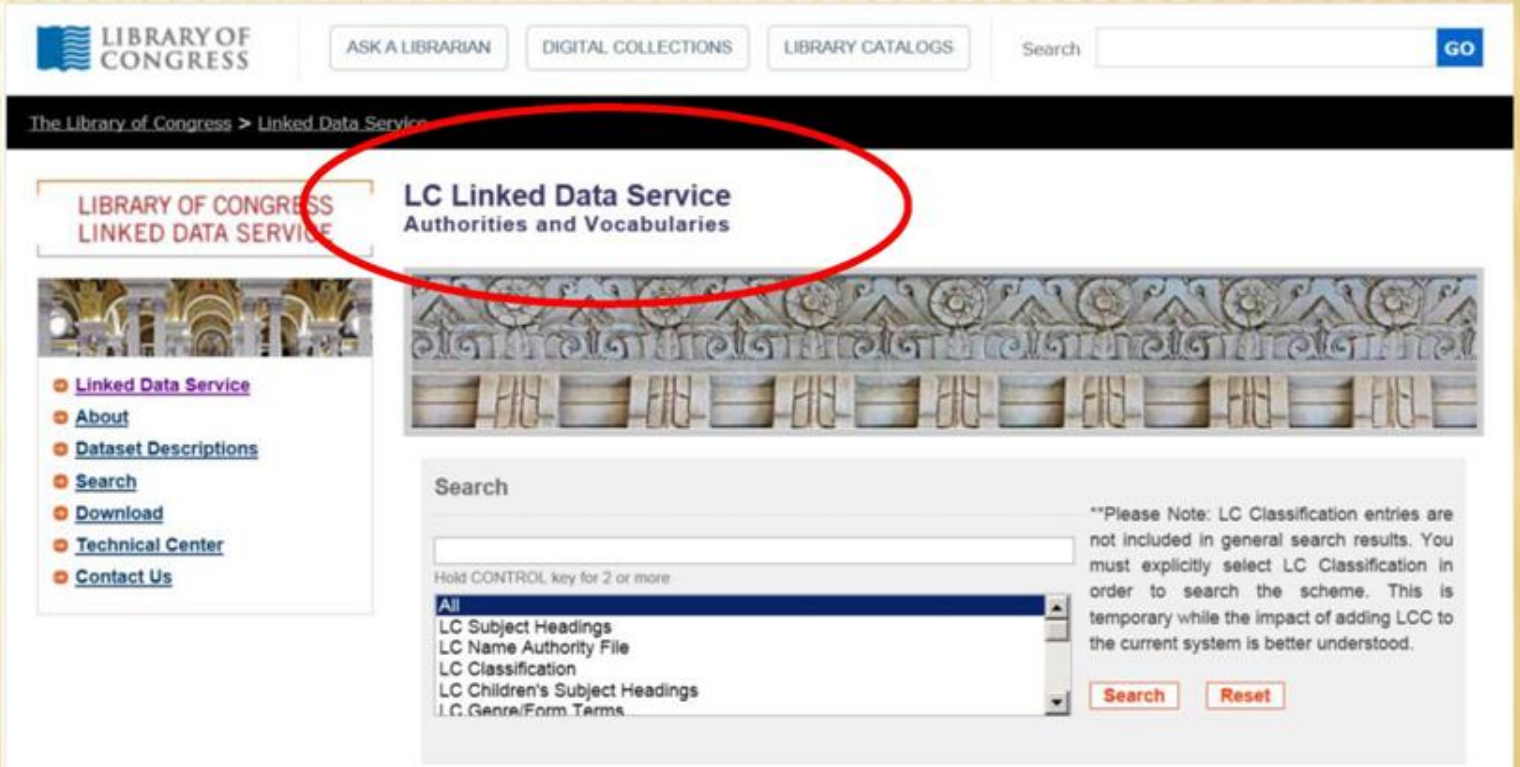


Библиотека Конгресса США

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.



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.

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.