

Semantic Web, связанные данные и библиотеки

О. Н. Жлобинская

**Национальная служба развития системы
форматов RUSMARC**

**«ЛИБНЕТ-2012»
г.Звенигород**

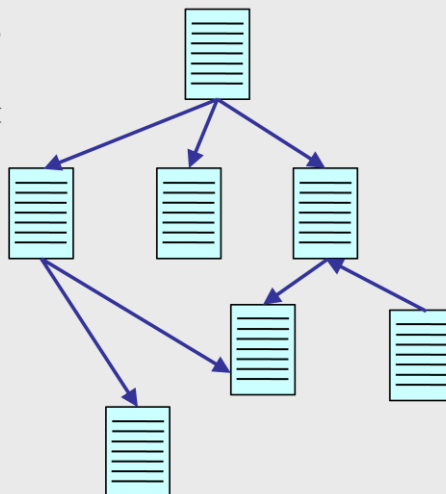
- ☐ **Что такое Semantic Web?**
- ☐ **Как работают связанные данные?**
- ☐ **Какие перспективы использования связанных данных открывает для библиотек?**
- ☐ **Что уже сделано?**
- ☐ **Что предстоит сделать?**

Современному пользователю в поисках информации удобнее обратиться к интернету, нежели к библиотечным каталогам (даже электронным) – для него это и удобнее, и быстрее. Библиотечные каталоги, несмотря на все богатство заключенной в них информации, остаются в большой степени изолированными от остального информационного поля и недоступными для поисковых машин в интернете. Принимая этот факт и понимая, что в современных условиях такое изолированное существование – путь в никуда, библиотечное сообщество в последние годы ищет пути, которые бы обеспечили интеграцию библиотечных данных в универсальное информационное пространство.

Наиболее перспективным в этом отношении представляется использование *Semantic Web* и *связанных данных*. Попробуем немного разобраться – почему это так важно? Что такое Semantic Web? Как работают связанные данные? Какие перспективы их использования открывает для библиотек? Что уже сделано в этом направлении и что предстоит сделать?

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.



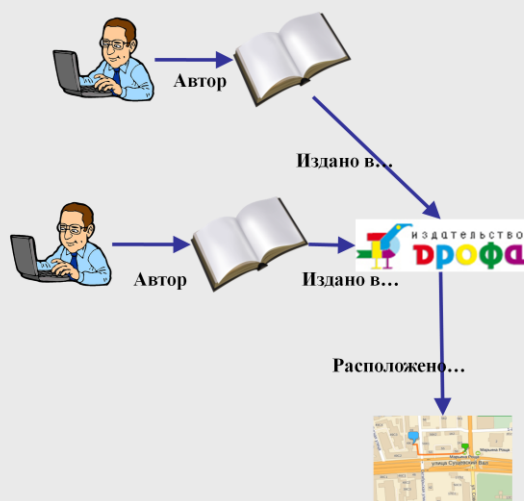
Чтобы понять, что такое Semantic Web, рассмотрим его в сравнении с привычной нам Всемирной паутиной – Web 2.0.

Традиционный Web – это сеть сайтов, использующих гиперссылки для перехода между страницами (документами). Для установки ссылок на веб-страницу, файл и т.д. используются адреса – универсальные идентификаторы ресурсов. Эти гиперссылки, по сути, и создают Всемирную Паутину.

Связи в традиционном вебе:

- Устанавливаются между документами, а точнее – между текстами документов, поэтому Web 2.0 иногда называют сетью документов.
- Односторонние – в документе, на который ссылается другой документ, нет никакой информации об этой ссылке; обнаружить ее можно, но только с помощью специальных инструментов.
- Неинформативные (говорят только о наличии связи, но не о ее характере – что означает сама связь); это может быть библиографическая ссылка, либо ссылка на источник, где можно получить дальнейшую информацию, и т.д.

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 и 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.

Semantic Web, или Web 3.0 – следующий этап развития Интернета, новый уровень представлений данных во Всемирной паутине. В Semantic Web связи устанавливаются не между текстами, а между данными, а точнее – между произвольными сущностями. Это могут быть не только веб-страницы, но и любые объекты реального мира (люди, города, художественные произведения), и даже абстрактные понятия (например, свойства «имя», «должность», «цвет»). В отличие от традиционного веба («веба документов»), Web 3.0 называют *сетью данных*.

Для установления ссылок также используются универсальные идентификаторы ресурсов (URI), однако в отличие от обычного веба, в Semantic Web URI используются также для именования объектов и абстрактных понятий.

Идентификаторы обязаны быть уникальными, это позволяет называть одни и те же объекты и обращаться к ним в разных местах в семантической паутине.

Связанные данные

Связанные данные – термин, используемый для обозначения методов публикации, обмена и установления связей между отдельными данными и знаниями в Semantic Web с использованием URI и RDF (*Википедия*)



1. В качестве имен объектов (в том числе и абстрактных понятий) используются уникальные идентификаторы (URI)
2. Для того, чтобы обеспечить возможность различать имена и обращаться к объектам, используется HTTP
3. В URI следует представлять полезную информацию с использованием стандартных языков (прежде всего RDF) - основной формат - RDF/XML, но возможны и другие (например, N3, или Turtle).
4. Обязательное условия – наличие ссылок на другие URI с целью раскрытия дополнительной информации о сущности

Тим Бернерс-Ли , 2006 г.

Итак, ключевым в понятии Semantic Web является не метод размещения данных в Web, а установление связей между данными таким образом, что данные могут быть поняты и использованы компьютером и человеком.

Основные принципы Semantic Web и связанных данных сформулировал Тим Бернерс Ли в 2006 г.:

1. В качестве имен объектов (в том числе и абстрактных понятий) используются уникальные идентификаторы (URI)
2. Для того, чтобы обеспечить возможность различать имена и обращаться к объектам, используется HTTP
3. В URI следует представлять полезную информацию с использованием стандартных языков (прежде всего RDF).
Основной формат - RDF/XML, но возможны и другие (например, N3, или Turtle).
4. Обязательное условия – наличие ссылок на другие URI с целью раскрытия дополнительной информации о сущности

Существует множество определений связанных данных; одно из них (из Википедии) представлено на слайде:

Связанные данные – термин, используемый для обозначения методов публикации, обмена и установления связей между отдельными данными и знаниями в Semantic Web с использованием URI и RDF.

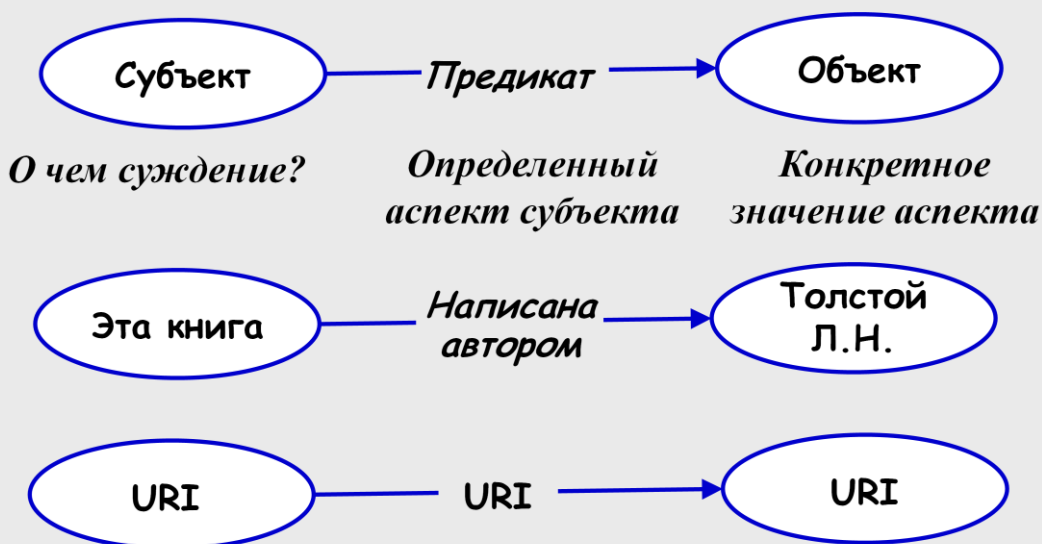
Еще несколько определений:

Связанные данные – технология установления ссылок между сущностями (данными, понятиями, документами) в Web, используя URI для идентификации, RDF для описания и HTTP для публикации этих сущностей и связей между ними, таким образом, что они могут быть поняты и использованы компьютером и человеком.

Связанные данные – это данные, опубликованные в соответствии с принципами, обеспечивающими установление связей между наборами данных, наборами элементов и словами значений.

Во всех определениях ключевыми являются (а) собственно связи, (б) URI - универсальные идентификаторы ресурсов и (в) RDF – Resource Description Framework (Среда описания ресурсов).

RDF (Resource Description Framework)



RDF (Resource Description Framework – среда описания ресурса) – это графовая модель описания связанных данных, которая позволяет технологии Semantic Web интерпретировать информацию, представленную в Web. RDF определяет связи между сущностями; эти связи могут использоваться для навигации, или интеграции информации из разных источников.

В RDF данные выражаются в виде отдельных суждений, представленных тройками (или триплетами) в форме *Субъект* (идентифицирует, о чем это суждение) – *Предикат* (связывает субъект и объект, идентифицирует определенный аспект описываемого субъекта) – *Объект* (идентифицирует конкретное значение этого аспекта). Субъект и объект – экземпляры класса RDF, а предикат представлен как RDF-свойство.

Тройка может быть представлена в виде простого графа, в котором субъект и объект – это узлы, а предикат изображается дугой или иной соединительной линией, направленной от субъекта к объекту.

Например: *Эта книга – написана автором – Л.Н. Толстой*

Для того, чтобы обеспечить однозначную компьютерную обработку таких суждений и строить графы связанных данных из отдельных троек, все сущности в RDF представляются не наименованиями, а уникальными идентификаторами (URI – универсальными идентификаторами ресурса). Внешне URI может выглядеть как привычный электронный адрес. Но в отличие от URL, URI не ведет к ресурсу - это лишь идентифицирующая метка.

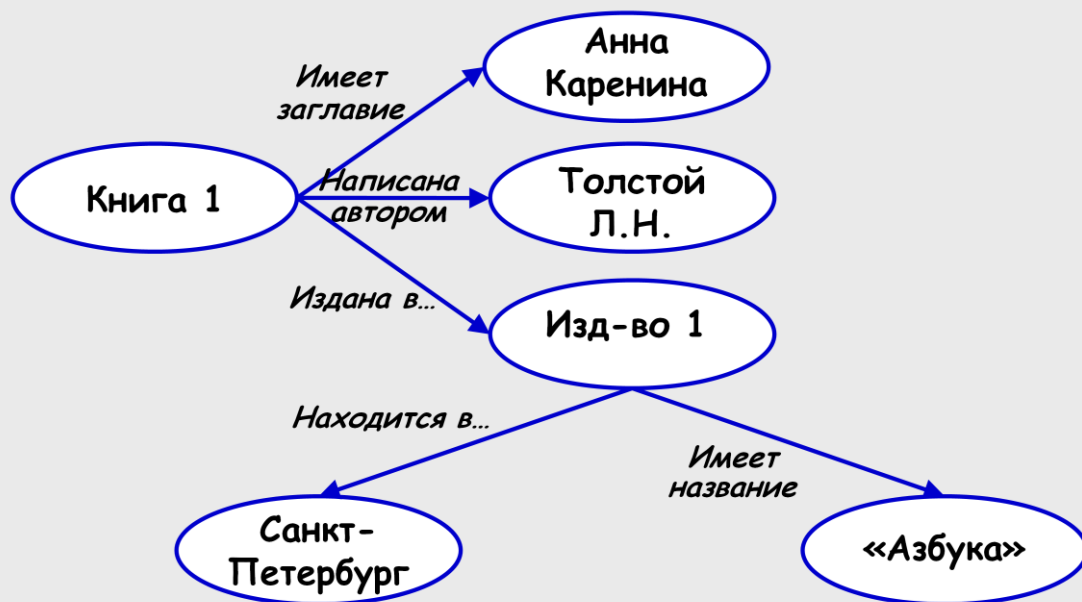
RDF (Resource Description Framework)



Цепочки троек объединяются в граф: так, в нашем примере субъект первой тройки (*Книга 1*) является также субъектом троек 2 и 3 (*Книга 1 – написана автором – Толстой Л.Н.* и *Книга 1 – издана в... – Издательство 1*). Объект тройки 3 (*Издательство 1*) является субъектом троек 4 и 5 (*Издательство 1 – имеет название – «Азбука»* и *Издательство 1 – находится в – Санкт-Петербург*).

Тройки с совпадающими субъектами или объектами объединяются в граф (см. следующий слайд):

RDF (Resource Description Framework)



Используя именованные связи, программа может проследить всю цепочку.

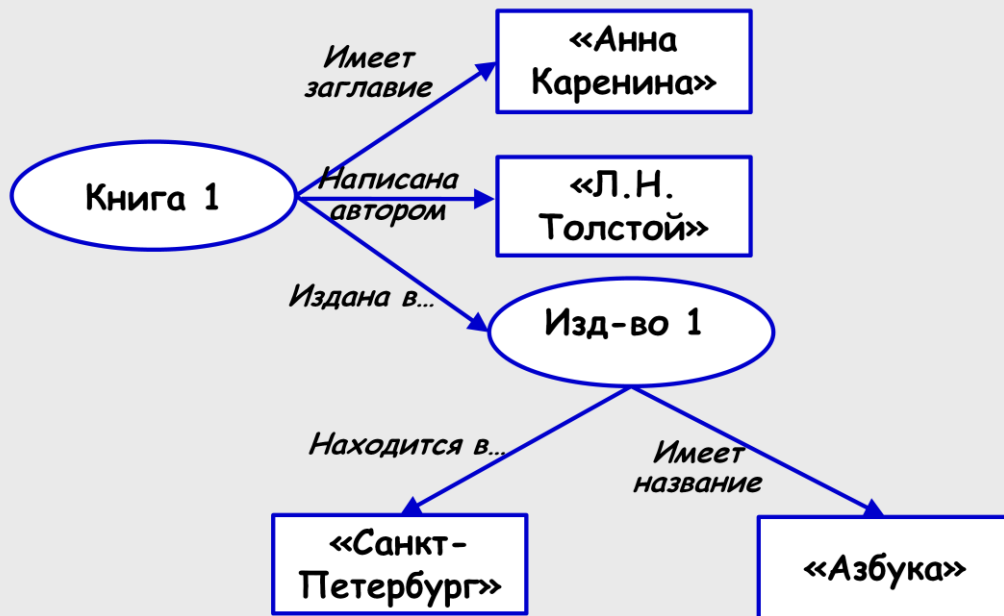
Очевидно, что информативность (а следовательно, и полезность) данных увеличивается по мере увеличения количества связей с другими данными.

Субъект и предикат всегда выражены только идентификатором. Объект может также быть представлен идентификатором, либо литералом – текстовой строкой. Это может быть имя лица, заглавие ресурса, наименование организации, предметная рубрика.

На схемах принято следующее обозначение: субъект и объект, имеющие URI, представляются овалами; объект-литерал – прямоугольником. Предикат представляется однонаправленной дугой, указывающей на объект.

То есть RDF-граф в нашем примере представляется в виде (см. следующий слайд):

RDF (Resource Description Framework)



Литералы всегда завершают цепочку и используются для человеко-читаемого вывода (дисплейного представления). То есть читатель видит только то, что находится на конце цепочки связанных данных (заглавие, имя автора, географическое название и т.д.); все промежуточные звенья этой цепочки от него, как правило, скрыты, и внешне он может даже и не знать, что имеет дело со связанными данными.

Как видим, подход RDF очень отличается от традиционной каталожной записи (например, в формате RUSMARC). В привычной нам записи различные аспекты ресурса объединяются в одну цепочку (запись) с помощью особого синтаксиса – тегов полей, индикаторов, подполей. В RDF данные разделены на отдельные суждения, которые могут обрабатываться независимо друг от друга; важно отметить, что *могут использоваться суждения, или тройки, из разных источников и разных схем метаданных*.

Идентификаторы и пространства имен

- Структура идентификатора:
 - базовый элемент (домен)
`http://iflastandards.info/ns/isbd/elements`
 - локальная часть – уникальный идентифицирующий суффикс
`P1001`
- Базовый элемент определяет пространство имен (namespace).
- Пространство имен – множество URI с базовой информацией о соответствующих классах и свойствах, опубликованное и управляемое в едином контексте (по сути - множество URI с одним и тем же доменом).

Второе условие связанных данных – использование универсальных идентификаторов ресурсов, которые необходимы для однозначной идентификации каждой части тройки.

Идентификатор может представлять собой любую комбинацию любых символов (букв и цифр). Никаких ограничений на форму идентификатора нет, кроме одного – он должен быть уникальным. Внутреннего значения идентификатор также не несет, это лишь «ярлык», идентифицирующий сущность.

Внешне может выглядеть как URL, но это не URL – идентификатор не ведет к веб-странице.

Как правило, для записи универсальных идентификаторов ресурсов используется `http`. Использование `http` – не единственный способ, но удобен тем, что использует инфраструктура web.

URI может строиться из базового элемента (домена, или общей части для схемы или словаря) и уникального идентифицирующего суффикса, например:

- `http://iflastandards.info/ns/isbd/elements`
- `P1001`

Базовый элемент, или префикс в именах идентификаторов (по сути - множество URI с одним и тем же доменом), определяет *пространство имен (namespace)*. По сути, пространство имен – это множество URI с базовой информацией о соответствующих классах и свойствах, опубликованное и управляемое в едином контексте. Это некий абстрактный контейнер для множества терминов (имен, слов и т.д.), который создает контекст для использования терминов и позволяет различать термины, относящиеся к разным пространствам имен.

Для компактности пространство имен в записи идентификаторов может быть представлено так называемым “qname” – сокращенной формой, например: *isbd* вместо <http://iflastandards.info/ns/isbd/elements/>.

Для пользователя может выводиться сокращенная форма, а при компьютерной обработке раскрывается в полную форму.

Связанные данные. Новые возможности для библиотек.

- ☐ **Трансформация библиотечных данных в форму, обеспечивающую их интеграцию в Web**
- ☐ **Библиотечные метаданные становятся доступными для поисковых машин в Web**
- ☐ **Более эффективное хранение метаданных**
- ☐ **Совместное использование информации**
- ☐ **Возможность многократного использования данных**
- ☐ **Более гибкие возможности при описании ресурсов**
- ☐ **Более гибкий вывод результатов для пользователя**

Что нам дает использование связанных данных?

1. Трансформация библиотечных данных в форму, обеспечивающую их интеграцию в Web. Разработчики не будут обязаны работать с чисто библиотечными форматами данных (такими как MARC), требующими особого ПО. Метод связанных данных предполагает публикацию данных в Web в форме, которая является общепонятной.
2. Библиотечные метаданные становятся доступными для поисковых машин в Web – ссылки к библиотечным коллекциям могут устанавливаться извне и в форме, понятной для поисковых машин.
3. Более эффективное хранение метаданных. Хранится не запись, а отдельные ее компоненты – тройки; запись собирается «на лету». Любое изменение сразу отражается в итоговой записи и становится доступно всем пользователям (можно провести аналогию с авторитетными записями).
4. Возможность совместного использования информации – не только библиотеками, но и другими сообществами
5. Возможность многократного использования данных
6. Более гибкие возможности при описании ресурсов (расширения и дополнения информации)
7. Более гибкий вывод результатов для пользователя - детальность описания (на уровне элементарных утверждений-троек) позволяет строить гибкий вывод метаданных в соответствии с практикой библиотеки и потребностями пользователя.

Внешне пользователь может и не знать, что использует связанные. Однако, поскольку структурированные данные станут более насыщены взаимосвязями, пользователь может заметить улучшение возможностей поиска и использования данных. Навигация по ресурсам разных библиотек (и-не библиотечным ресурсам) переходит на более высокий уровень.

Библиотечные ресурсы как связанные данные

- 1. Наборы данных – коллекции структурированных метаданных о библиографических ресурсах, связанных с ними лицах и организациях, предметах и т.д.
– библиотечные каталоги**
- 2. Значения или словари значений – контролируемые перечни терминов
– тезаурусы, перечни кодов, схемы классификации, списки предметных рубрик, авторитетные файлы и другие системы организации знаний**
- 3. Элементы метаданных или словари элементов метаданных – схемы и форматы метаданных (в т.ч. ISBD, FRBR, MARC-форматы)**

W3C. Library Linked Data Incubator Group (LLDXG)

В том, чтобы библиотеки активнее использовали Semantic Web и связанные данные, проявляют заинтересованность как библиотеки, так и сообщество Web.

Так, в отчете Рабочей группы Консорциума Всемирной паутины Library Linked Data Incubator Group (LLDXG), работавшей в 2010-2011 гг., отмечено, что библиотеки накопили огромный опыт в области генерации метаданных, поиска, разработки онтологий. Учитывая это, они могут и должны внести свой вклад в развитие и внедрение связанных данных.

Выделены три категории данных, которые представляют интерес с точки зрения их использования в виде связанных данных.

1. Наборы данных – коллекции структурированных метаданных о библиографических ресурсах (библиотечные каталоги)
2. Словари значений – контролируемые перечни терминов, которые используются в качестве значений в библиографических записях (тезаурусы, перечни кодов, схемы классификации, списки предметных рубрик, авторитетные файлы и другие системы организации знаний)
3. Элементы метаданных или наборы элементов метаданных – схемы и форматы метаданных (в т.ч. ISBD, FRBR, MARC-форматы)

Библиотечные связанные данные. Что сделано? Контролируемые словари

SKOS (Simple Knowledge Organization System – Простая система организации знаний)

- **Словари RDA**
- **LCSH (Library of Congress Subject Headings)**
- **TGM (Thesaurus for Graphic Materials)**
- **FAST (Faceted Application of Subject Terminology)**
- **MESH (Medical Subject Headings)**
- **Form and genre headings for fiction and drama**
- **SWD (немецкие предметные термины)**
- **RAMEAU (Французские предметные рубрики)**
- **UDC summary**
- **VIAF (Virtual International Authority File)**

В настоящее время в виде связанных данных публикуются в основном словари значений и словари элементов метаданных.

Словари значений

Термины из контролируемых словарей (ПР, классификационные индексы и т.д.) могут быть представлены в RDF с помощью SKOS. (Simple Knowledge Organization System – «простая система организации знаний») основана на RDF и определяет модель представления базовой структуры и содержания тезаурусов, рубрикаторов и т.д.

SKOS поддерживает простые связи между терминами - иерархические, ассоциативные, эквивалентности; как правило, этих связей достаточно для представления внутренней структуры системы терминов.

- Словари RDA (OMR, 2011) – плоские словари (перечни терминов) без иерархических связей.
- LCSH (Library of Congress Subject Headings) (<http://id.loc.gov/>)
- FAST (Faceted Application of Subject Terminology)
- MESH (Medical Subject Headings)
- Form and genre headings for fiction and drama
- TGM (Thesaurus for Graphic Materials)
- SWD (Немецкие предметные термины)
- RAMEAU (Французские предметные рубрики)
- UDC summary (около 2400 классов УДК на 40 языках (<http://udcdata.info>))
- VIAF (Virtual International Authority File) – множество связанных контролируемых словарей (авторитетные записи на имена лиц) национальных библиографирующих агентств.

Этот список постоянно пополняется.

Библиотечные связанные данные. Что сделано?

Элементы метаданных

1. Семейство моделей FRBR

2007 Проект “Пространство имен для FRBR” в рамках рабочей группы IFLA FRBR Review Group

Цель проекта – определить соответствующие пространства имен для модели FRBR в RDF.

2011 Модели FRBR, FRAD и FRSAD опубликованы в OMR (Open Metadata Registry).

<http://metadataregistry.org/schema/list.html>

Наборы элементов метаданных в виде связанных данных

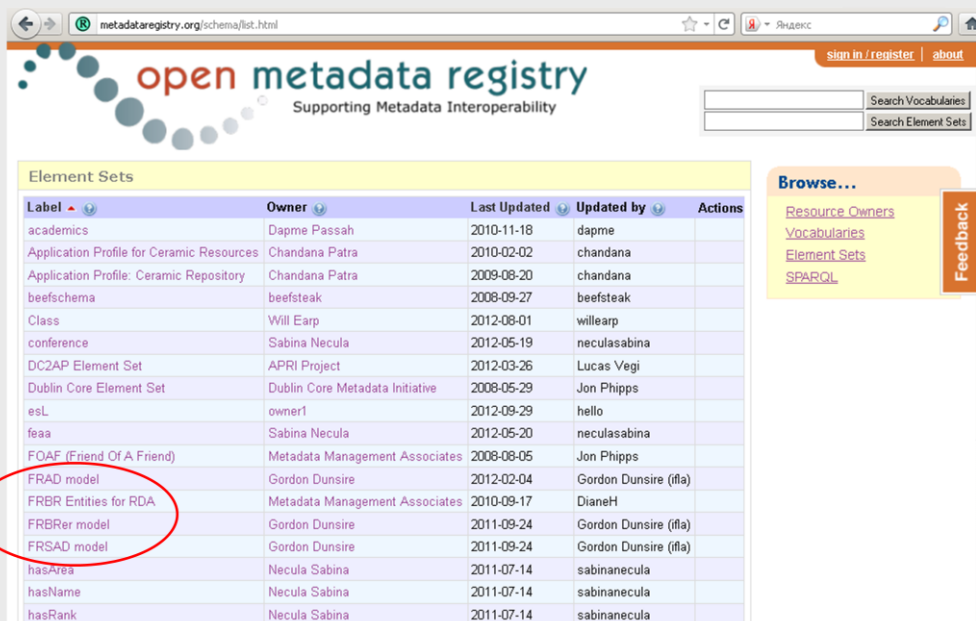
1. FRBR

2007 – В рамках рабочей группы IFLA (FRBR Review Group) – инициирован проект “Пространство имен для FRBR”, цель которого – определить соответствующие пространства имен для модели FRBR (сущность – отношение) в RDF.

2011 – FRBR, FRAD и FRSAD опубликованы в OMR (Open Metadata Registry – Открытый регистр метаданных).

Библиотечные связанные данные. Что сделано?

1. Семейство моделей FRBR



Label	Owner	Last Updated	Updated by	Actions
academics	Dapme Passah	2010-11-18	dapme	
Application Profile for Ceramic Resources	Chandana Patra	2010-02-02	chandana	
Application Profile: Ceramic Repository	Chandana Patra	2009-08-20	chandana	
beefschema	beefsteak	2008-09-27	beefsteak	
Class	Will Earp	2012-08-01	willearp	
conference	Sabina Nacula	2012-05-19	neculasabina	
DC2AP Element Set	APRI Project	2012-03-26	Lucas Vegi	
Dublin Core Element Set	Dublin Core Metadata Initiative	2008-05-29	Jon Phipps	
esL	owner1	2012-09-29	hello	
feaa	Sabina Nacula	2012-05-20	neculasabina	
FOAF (Friend Of A Friend)	Metadata Management Associates	2008-08-05	Jon Phipps	
FRAD model	Gordon Dunsire	2012-02-04	Gordon Dunsire (ifla)	
FRBR Entities for RDA	Metadata Management Associates	2010-09-17	DianeH	
FRBRer model	Gordon Dunsire	2011-09-24	Gordon Dunsire (ifla)	
FRSAD model	Gordon Dunsire	2011-09-24	Gordon Dunsire (ifla)	
hasArea	Nacula Sabina	2011-07-14	sabinanacula	
hasName	Nacula Sabina	2011-07-14	sabinanacula	
hasRank	Nacula Sabina	2011-07-14	sabinanacula	

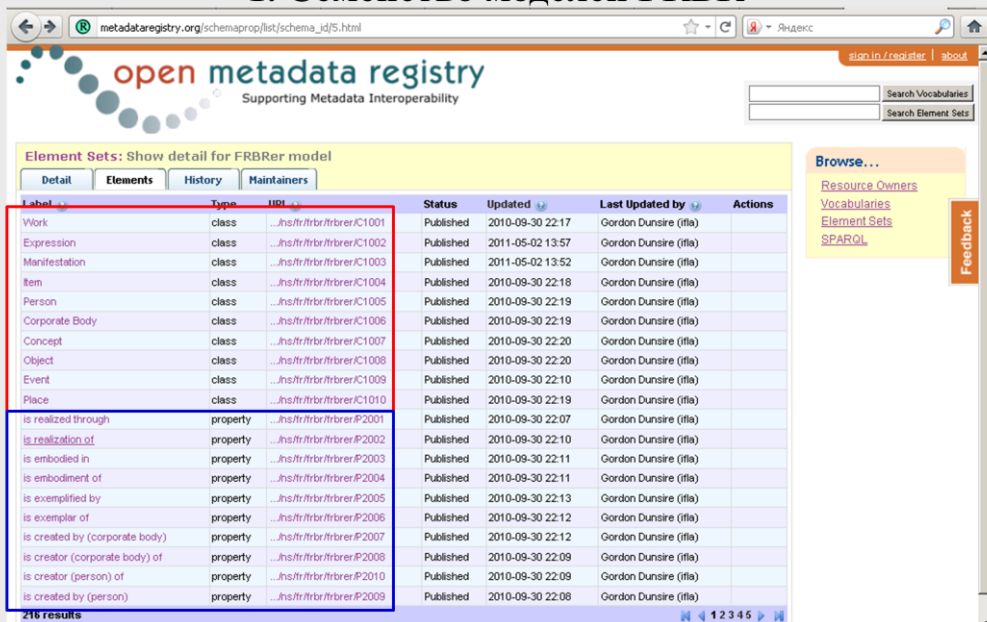
Каждая модель (FRBR, FRAD и FRSAD) имеет собственное пространство имен; продолжается работа над консолидированной моделью.

На слайде – фрагмент перечня пространства имен из OMR

(<http://metadataregistry.org/schema/list.html>)

Библиотечные связанные данные. Что сделано?

1. Семейство моделей FRBR



The screenshot shows the 'open metadata registry' website. The main heading is 'Element Sets: Show detail for FRBRer model'. Below this, there are tabs for 'Detail', 'Elements', 'History', and 'Maintainers'. The 'Elements' tab is selected, displaying a table of FRBR elements. The table has columns: Label, Type, URI, Status, Updated, Last Updated by, and Actions. The elements are listed in two groups: 'Work' (classes) and 'is realized through' (properties). The 'Work' group includes 'Work', 'Expression', 'Manifestation', 'Item', 'Person', 'Corporate Body', 'Concept', 'Object', 'Event', and 'Place'. The 'is realized through' group includes 'is realized through', 'is realization of', 'is embodied in', 'is embodiment of', 'is exemplified by', 'is exemplar of', 'is created by (corporate body)', 'is creator (corporate body) of', 'is creator (person) of', and 'is created by (person)'. The table shows that all elements are 'Published' and were last updated by 'Gordon Dunsire (ifla)'.

Label	Type	URI	Status	Updated	Last Updated by	Actions
Work	class	.../frbr/irbrc1001	Published	2010-09-30 22:17	Gordon Dunsire (ifla)	
Expression	class	.../frbr/irbrc1002	Published	2011-05-02 13:57	Gordon Dunsire (ifla)	
Manifestation	class	.../frbr/irbrc1003	Published	2011-05-02 13:52	Gordon Dunsire (ifla)	
Item	class	.../frbr/irbrc1004	Published	2010-09-30 22:18	Gordon Dunsire (ifla)	
Person	class	.../frbr/irbrc1005	Published	2010-09-30 22:19	Gordon Dunsire (ifla)	
Corporate Body	class	.../frbr/irbrc1006	Published	2010-09-30 22:19	Gordon Dunsire (ifla)	
Concept	class	.../frbr/irbrc1007	Published	2010-09-30 22:20	Gordon Dunsire (ifla)	
Object	class	.../frbr/irbrc1008	Published	2010-09-30 22:20	Gordon Dunsire (ifla)	
Event	class	.../frbr/irbrc1009	Published	2010-09-30 22:10	Gordon Dunsire (ifla)	
Place	class	.../frbr/irbrc1010	Published	2010-09-30 22:19	Gordon Dunsire (ifla)	
is realized through	property	.../frbr/irbrc1001	Published	2010-09-30 22:07	Gordon Dunsire (ifla)	
is realization of	property	.../frbr/irbrc1002	Published	2010-09-30 22:10	Gordon Dunsire (ifla)	
is embodied in	property	.../frbr/irbrc1003	Published	2010-09-30 22:11	Gordon Dunsire (ifla)	
is embodiment of	property	.../frbr/irbrc1004	Published	2010-09-30 22:11	Gordon Dunsire (ifla)	
is exemplified by	property	.../frbr/irbrc1005	Published	2010-09-30 22:13	Gordon Dunsire (ifla)	
is exemplar of	property	.../frbr/irbrc1006	Published	2010-09-30 22:12	Gordon Dunsire (ifla)	
is created by (corporate body)	property	.../frbr/irbrc1007	Published	2010-09-30 22:12	Gordon Dunsire (ifla)	
is creator (corporate body) of	property	.../frbr/irbrc1008	Published	2010-09-30 22:09	Gordon Dunsire (ifla)	
is creator (person) of	property	.../frbr/irbrc1009	Published	2010-09-30 22:09	Gordon Dunsire (ifla)	
is created by (person)	property	.../frbr/irbrc1010	Published	2010-09-30 22:08	Gordon Dunsire (ifla)	

Сущности, или объекты FRBR (произведение, выражение, воплощение, физическая единица, лицо, организация, концепция, предмет, событие, место) в пространстве имен определены как классы, а атрибуты (реализовано в..., является реализацией..., воплощено в..., является воплощением..., и т.д.) – как свойства.

Библиотечные связанные данные. Что сделано?

Элементы метаданных

2. ISBD

- 2008** Рабочая группа IFLA ISBD/XML (ISBD/XML Study Group); одна из задач - пересмотр понятий ISBD и самого стандарта с учетом возможности применения современных web-технологий, в частности, построение схемы ISBD RDF/XML, и создание на основе стандарта современного инструмента, открытого для технологий и служб Semantic Web.
- 2011** Определен набор элементов, а также словари терминов формы содержания и типа средства (Область 0), опубликовано пространство имен в OMR (Open Metadata Registry).

<http://metadataregistry.org/schema/list.html>

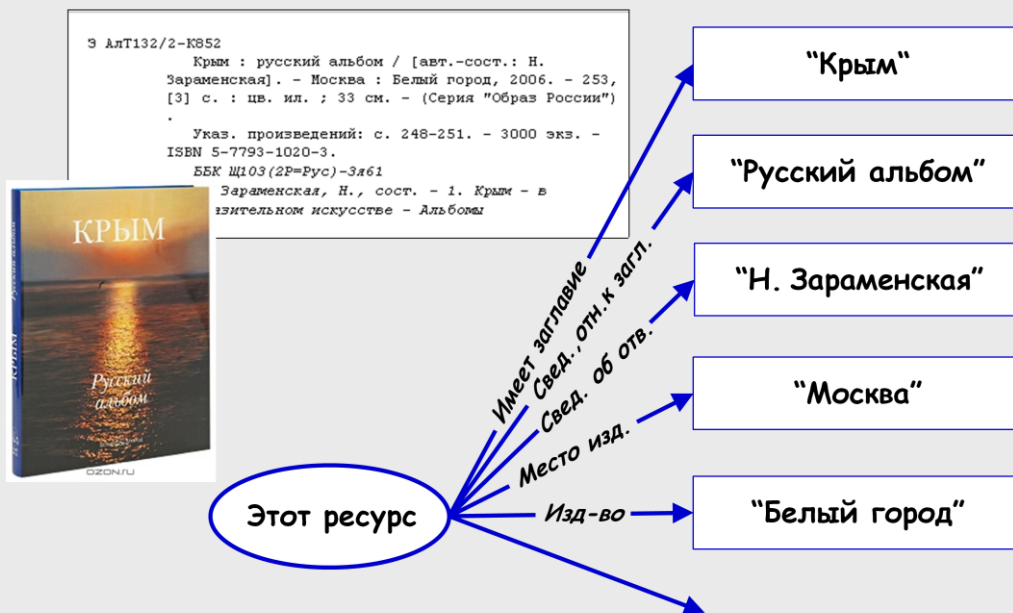
2. ISBD.

2008 – Создана Рабочая группа IFLA ISBD/XML (ISBD/XML Study Group), среди задач рабочей группы – пересмотр понятий ISBD и самого стандарта с учетом возможности применения современных web-технологий, в частности, построение схемы ISBD RDF/XML, и создание на основе стандарта современного инструмента, открытого для технологий и служб Semantic Web.

2011 – определен набор элементов, а также словари терминов формы содержания и типа средства (Область 0), опубликовано пространство имен в OMR.

Библиотечные связанные данные. Что сделано?

2. ISBD



В отличие от концептуальной модели FRBR, которая построена на методологии E-R (сущность-отношение), по своей природе более приспособленной для представления в RDF, ISBD представляет собой стандарт данных, и в этом случае используется другой подход.

RDF определяет элементы метаданных как классы или свойства. Класс – это сущность, которая описывается в утверждении, свойство – это некий аспект этой сущности или отношение между разными сущностями.

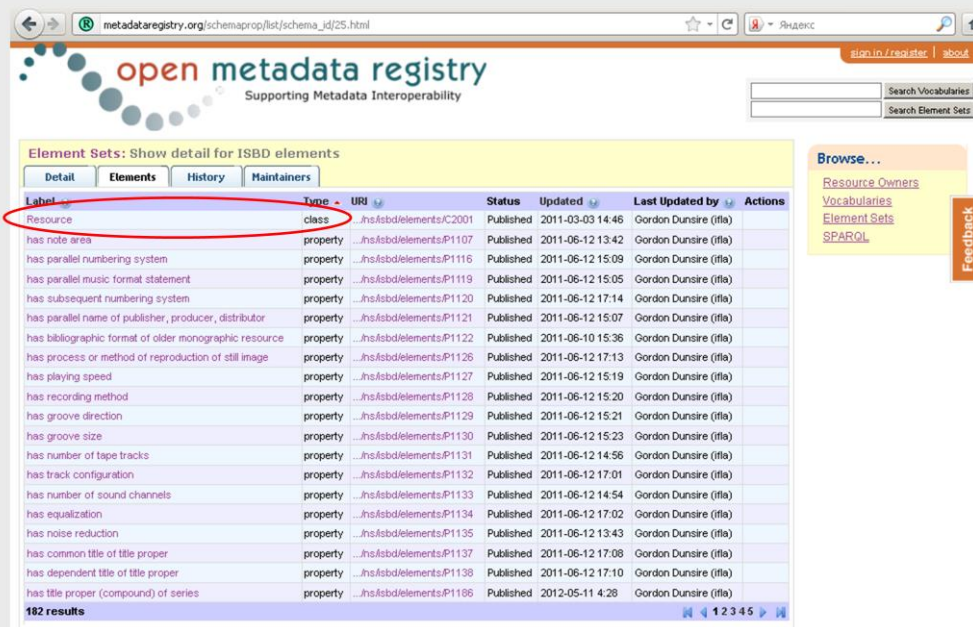
ISBD описывает единственную сущность – библиографический ресурс, поэтому в RDF ресурс определен как класс. Все атрибуты ISBD представляют собой некие аспекты, или характеристики описываемого ресурса, поэтому каждый атрибут представлен в виде свойства.

Например, атрибут "основное заглавие" представлен свойством "имеет основное заглавие". Такой принцип именования свойств (имеет...) используется в RDF-представлении ISBD для читабельности и указания «направленности» свойства.

Таким образом, запись ISBD представляется в виде, показанном на слайде.

Библиотечные связанные данные. Что сделано?

2. ISBD



metadateregistry.org/schemaprop/list/schema_id/25.html

open metadata registry
Supporting Metadata Interoperability

Element Sets: Show detail for ISBD elements

Label	Type	URI	Status	Updated	Last Updated by	Actions
Resource	class	.../isbd/elements/C2001	Published	2011-03-03 14:46	Gordon Dunsire (ifla)	
has note area	property	.../isbd/elements/P1107	Published	2011-06-12 13:42	Gordon Dunsire (ifla)	
has parallel numbering system	property	.../isbd/elements/P1116	Published	2011-06-12 15:09	Gordon Dunsire (ifla)	
has parallel music format statement	property	.../isbd/elements/P1119	Published	2011-06-12 15:05	Gordon Dunsire (ifla)	
has subsequent numbering system	property	.../isbd/elements/P1120	Published	2011-06-12 17:14	Gordon Dunsire (ifla)	
has parallel name of publisher, producer, distributor	property	.../isbd/elements/P1121	Published	2011-06-12 15:07	Gordon Dunsire (ifla)	
has bibliographic format of older monographic resource	property	.../isbd/elements/P1122	Published	2011-06-10 15:36	Gordon Dunsire (ifla)	
has process or method of reproduction of still image	property	.../isbd/elements/P1126	Published	2011-06-12 17:13	Gordon Dunsire (ifla)	
has playing speed	property	.../isbd/elements/P1127	Published	2011-06-12 15:19	Gordon Dunsire (ifla)	
has recording method	property	.../isbd/elements/P1128	Published	2011-06-12 15:20	Gordon Dunsire (ifla)	
has groove direction	property	.../isbd/elements/P1129	Published	2011-06-12 15:21	Gordon Dunsire (ifla)	
has groove size	property	.../isbd/elements/P1130	Published	2011-06-12 15:23	Gordon Dunsire (ifla)	
has number of tape tracks	property	.../isbd/elements/P1131	Published	2011-06-12 14:56	Gordon Dunsire (ifla)	
has track configuration	property	.../isbd/elements/P1132	Published	2011-06-12 17:01	Gordon Dunsire (ifla)	
has number of sound channels	property	.../isbd/elements/P1133	Published	2011-06-12 14:54	Gordon Dunsire (ifla)	
has equalization	property	.../isbd/elements/P1134	Published	2011-06-12 17:02	Gordon Dunsire (ifla)	
has noise reduction	property	.../isbd/elements/P1135	Published	2011-06-12 13:43	Gordon Dunsire (ifla)	
has common title of title proper	property	.../isbd/elements/P1137	Published	2011-06-12 17:08	Gordon Dunsire (ifla)	
has dependent title of title proper	property	.../isbd/elements/P1138	Published	2011-06-12 17:10	Gordon Dunsire (ifla)	
has title proper (compound) of series	property	.../isbd/elements/P1186	Published	2012-05-11 4:28	Gordon Dunsire (ifla)	

182 results

Browse...
Resource Owners
Vocabularies
Element Sets
SPARQL

Feedback

Пространство имен ISBD опубликовано в OMR. Единственный RDF-класс – это ресурс (описываемый в ISBD-записи); все атрибуты ISBD представлены как RDF-свойства (имеет область примечаний..., имеет параллельную систему нумерации, и т.д.).

Библиотечные связанные данные. Что сделано? IFLA Namespaces Task Group

2009 Рабочая группа IFLA по пространствам имен (IFLA Namespaces Task Group) под эгидой Секции классификации и индексирования.

- **Мониторинг разработки пространств имен IFLA
- FRBR, FRAD, FRSAD, ISBD, MulDiCat**
- **Разработка соответствия / связей между пространствами имен IFLA**
- **Разработка связей с другими пространствами имен
- Dublin Core, MARC21, RDA**
- **Исследование пространств имен на предмет взаимодействия между схемами и данными смежных областей (архивы, библиотеки, музеи и т.д.)**
- **Разработка Руководства по переводу пространств имен**
- **Разработка Руководства по использованию пространств имен IFLA**

2009 – создана Рабочая группа IFLA по пространствам имен (IFLA Namespaces Task Group) под эгидой Секции классификации и индексирования, с участием представителей других секций (в т.ч. библиографии, каталогизации, информационных технологий и др.). Цель создания Рабочей группы – разработка административной и технической инфраструктуры для поддержки инициатив, связанных с представлением стандартов IFLA в формате, обеспечивающем их использование в Semantic Web.

Рабочая группа решает задачи:

- Мониторинг разработки пространств имен IFLA (FRBR, FRAD, FRSAD, ISBD)
- Разработка соответствия / связей между пространствами имен IFLA и с другими пространствами имен (Dublin Core, MARC21, RDA)
- Исследование пространств имен на предмет взаимодействия между схемами и данными смежных областей (архивы, библиотеки, музеи и т.д.)
- Разработка Руководства по переводу пространств имен (д.б. опубликовано до конца 2012 г.)
- Разработка Руководства по использованию пространств имен IFLA

Библиотечные связанные данные. Что сделано?

Элементы метаданных

3. UNIMARC и связанные данные

- 2011** **Предложен способ представления UNIMARC в RDF.**
Предлагается использовать подход, аналогичный ISBD и FRBR, с публикацией в OMR.

UNIMARC and linked data / Gordon Dunsire, Mirna Willer. Presented to meeting 187: Advancing UNIMARC: alignment and innovation, IFLA UNIMARC Programme (UNIMARC), World Library and Information Congress: 77th IFLA General Conference And Assembly 13-18 August 2011, San Juan, Puerto Rico.

UNIMARC и связанные данные / Gordon Dunsire, Mirna Willer.
<http://www.rusmarc.ru/publish/article.htm>

UNIMARC

2011 – на конференции ИФЛА в докладе М.Виллер и Г.Даншир предложен способ представления UNIMARC в RDF. Предлагается использовать подход, аналогичный ISBD и FRBR, с публикацией в OMR.

Библиотечные связанные данные. Что сделано?

3. UNIMARC и связанные данные

Пространства имен UNIMARC

UNIMARC/A – <http://iflastandards.info/ns/unimarc/unimarca/elements/>
(сокращенный вариант названия для вывода – unimarca)

UNIMARC/B – <http://iflastandards.info/ns/unimarc/unimarcb/elements/>
(сокращенный вариант названия для вывода – unimarcb)

Словари UNIMARC – для каждого словаря используется свой базовый домен.

Имя базового домена:

- общий базовый домен для словарей UNIMARC
(<http://iflastandards.info/ns/unimarc/terms/>)
- идентификатор конкретного словаря

Например:

<http://iflastandards.info/ns/unimarc/terms/graphicssmd>

– пространство имён для словаря терминов специфического обозначения материала в 116bba1.

Для каждого из форматов (UNIMARC/A и UNIMARC/B) требуется отдельное пространство имен, чтобы обеспечить возможность различить элементы данных (поля/подполя) с одинаковыми тегами.

для каждого из словарей UNIMARC используется свой базовый домен, аналогично практике ISBD. Имя базового домена строится из общего базового домена для словарей UNIMARC и идентификатора конкретного словаря. Идентификатор словаря, в свою очередь, представляет собой аббревиатуру названия словаря.

Коды и соответствующие величины и определения, используемые в Блоке кодированной информации для описания ресурса, наилучшим образом представляются как SKOS-словари, аналогично словарям Области 0 ISBD.

Могут использоваться и внешние словари - в случае, если набор кодированных величин формата UNIMARC в точности базируется на таком внешнем словаре. Например, в поле 101 используется трехсимвольный код, взятый из Приложения А, которая есть не что иное, как список формата MARC для языков. Этот список доступен как SKOS-словарь, который может использоваться непосредственно в любой тройке формата UNIMARC.

Библиотечные связанные данные. Что сделано?

3. UNIMARC и связанные данные

Использование существующих пространств имен

UNIMARC/B ⇔ ISBD

- (а) Использовать классы и свойства ISBD, а не создавать отдельные классы/свойства для UNIMARC
 - в том случае, если два стандарта функционируют как единое целое – т.е. если изменения, вносимые в один стандарт, предполагают параллельное внесение изменений в другой
- (б) Представить все элементы UNIMARC/B в отдельном пространстве имен UNIMARC, и связать их с эквивалентными классами и свойствами в пространстве имен ISBD
 - в том случае, если ISBD и UNIMARC в будущем будут, как и раньше, развиваться отдельно, даже при условии взаимосвязи между ними

UNIMARC/A ⇔ FRAD

Необходим более детальный анализ

Мы говорили, что связанные данные позволяют объединить данные из разных источников, использовать элементы и словари из уже существующих пространств имен – это позволяет сэкономить время и силы; упрощает разработку приложений для обработки метаданных, способствует использованию элементов из разных схем.

UNIMARC / B построен с учетом ISBD, для которого пространство имен уже опубликовано. Изменения ISBD, как правило, отражаются в формате, хотя это может занять какое-то время, и мы не можем гарантировать, что в любой момент времени UNIMARC полностью покрывает положения ISBD.

Поэтому для UNIMARC возможны варианты:

- (1) там, где это возможно, – использовать классы и свойства ISBD, а не создавать отдельные классы/свойства для UNIMARC.
- (2) представить все элементы UNIMARC/B в отдельном пространстве имен UNIMARC, и связать их с эквивалентными классами и свойствами в пространстве имен ISBD.

Выбор необходимо сделать ДО основных работ по разработке пространства имен UNIMARC.

Возможность использовать существующие пространства имен для UNIMARC/A требует дополнительного анализа.

Библиотечные связанные данные. Что сделано?

3. UNIMARC и связанные данные

Классы и свойства UNIMARC

Классы

UNIMARC/B – класс = ресурс

UNIMARC/A – ???

Свойства

Определяются сочетанием «**тег/индикатор/подполе**» :

URI свойства:

- Базовая часть (<http://iflastandards.info/ns/unimarc/unimarcb/elements/>)
- Локальная часть (например, P205bba)

Например:

<http://iflastandards.info/ns/unimarc/unimarcb/elements/P205bba>

1. Классы

Как и в случае ISBD, запись UNIMARC описывает единственную сущность – ресурс. Поскольку он в точности соответствует ресурсу ISBD – нет необходимости создавать для ресурса особый UNIMARC-класс, можно использовать класс из пространства имен ISBD. Классы для формата UNIMARC/Authorities требуют дальнейшего исследования, особенно в отношении FRAD/FRSAD.

2. Свойства

В общем случае RDF-свойства будут представлены сочетаниями тег/индикаторы/подполе. URI строится из базовой части (домен пространства имен) и локальной части, идентифицирующей конкретный элемент формата, например:

<http://iflastandards.info/ns/unimarc/unimarcb/elements/P205bba>.

Библиотечные связанные данные. Что сделано?

3. UNIMARC и связанные данные

Идентификация элементов формата

(а) Тег + Инд. 1 + Инд. 2 + код подполя

(для обозначения отсутствующего индикатора, а также вместо значка #, обозначающего пробел, используется символ «b»), например:

010##\$a (ISBN) ⇒ **010bba**

2001#\$a (Основное заглавие (значимое заглавие)) ⇒ **2001ba**

2000#\$a (Основное заглавие (незначимое заглавие)) ⇒ **2000ba**

(б) Блок кодированной информации: сокращенный код формируется путем добавления позиций символов, например:

100##\$a/8 (Тип даты публикации) ⇒ **100bba8**

100##\$a/17-19 (Код целевого назначения) ⇒ **100bba17-19**

100##\$a/34-35 (Графика заглавия) ⇒ **100bba34-35**

Для идентификации элементов формата UNIMARC (которые определяются тегам (поля), кодами подполей, индикаторами и позициями символов в блоке 1-- Блок кодированной информации) используются сокращенные коды, которые строятся следующим образом:

Тег + Инд. 1 + Инд. 2 + код подполя;

при этом для обозначения отсутствующего индикатора, а также вместо значка #, обозначающего пробел, используется символ «b».

Например:

010bba = ISBN

2001ba = Основное заглавие (значимое заглавие)

2000ba = Основное заглавие (незначимое заглавие)

Для Блока кодированной информации сокращенный код формируется путем добавления позиций символов, например:

100bba8 = Тип даты публикации

100bba17-19 = Код целевого назначения

100bba34-35 = Графика заглавия

Библиотечные связанные данные. Что сделано?

3. UNIMARC и связанные данные

2011 77-й Международный библиотечно-информационный конгресс IFLA

UNIMARC and linked data / Gordon Dunsire, Mirna Willer.

2012 78-й Международный библиотечно-информационный конгресс IFLA, Постоянный комитет по UNIMARC

Предложен проект (М.Виллер и Г.Даншир), включающий три этапа:

- **создание свойств формата UNIMARC (сначала для библиографических данных, потом – для авторитетных) и их загрузка в Регистр открытых метаданных – 2 года;**
- **построение семантических слоев, обеспечивающих установление связей с другими схемами;**
- **обеспечение интероперабельности с другими схемами.**

Такой подход предложен в 2011 г. на 77-м Конгрессе IFLA.

В 2012 г. на заседании Постоянного комитета IFLA по UNIMARC в рамках 78-го конгресса ИФЛА Мирна Виллер и Гордон Даншир в своем сообщении рассказали о результатах экспериментов по представлению формата MARC21 в RDF. В результате экспериментов были автоматически сгенерированы около 14000 свойств (properties), включая все возможные комбинации индикаторов, полей и подполей. Все они включены в Namespace Registry (Регистр пространств имен). Аналогичная операция может быть выполнена для UNIMARC, хотя для этого потребуются дополнительная интеллектуальная работа.

Мирна Виллер и Гордон Даншир выступили с инициативой по представлению формата UNIMARC в RDF.

Библиотечные связанные данные. Что сделано?

Библиографические записи

Представление записей в виде «троек»

- 1. Разбор библиографической записи на отдельные утверждения – «тройки»**
- 2. Создание URI для записи**
- 3. Замена идентификатора записи на URI**
- 4. Поиск URI для атрибутов**
- 5. Замена атрибутов на URI**
- 6. Поиск URI для значений**
- 7. Замена значений на URI**
- 8. Публикация троек**

⇒ Связанные данные

Наиболее перспективным для использования в Semantic Web представляются библиотечные каталоги, которые содержат огромный объем библиографических метаданных высокого качества, однако пока они в меньшей степени представлены в виде связанных данных.

Каталожные записи (которые могут быть представлены в разных форматах – UNIMARC, MARC21 и т.д.) могут быть также разобраны на «тройки». Этот процесс можно разложить на следующие этапы:

- 1. Разбор библиографической записи на отдельные утверждения-тройки»**
- 2. Создание URI для записи**
- 3. Замена идентификатора записи на URI**
- 4. Поиск URI для атрибутов**
- 5. Замена атрибутов на URI**
- 6. Поиск URI для значений**
- 7. Замена значений на URI**
- 8. Публикация троек и получение связанных данных**

Библиотечные связанные данные. Что сделано? Библиографические записи

- Британская национальная библиография
- Французская национальная библиография
- OCLC WorldCat

В настоящее время в виде связанных данных уже опубликованы ряд наборов библиографических записей (национальных библиографий и каталогов):

1. Британская национальная библиография

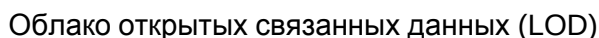
Используется модель связанных данных и различных пространств имен:

- значения из пространств имен MARC21 и VIAF (Виртуального международного авторитетного файла),
- элементы из ISBD и Dublin Core, а также
- локального пространства имен, разработанного специально для этой модели данных (для отражения отсутствующих элементов).

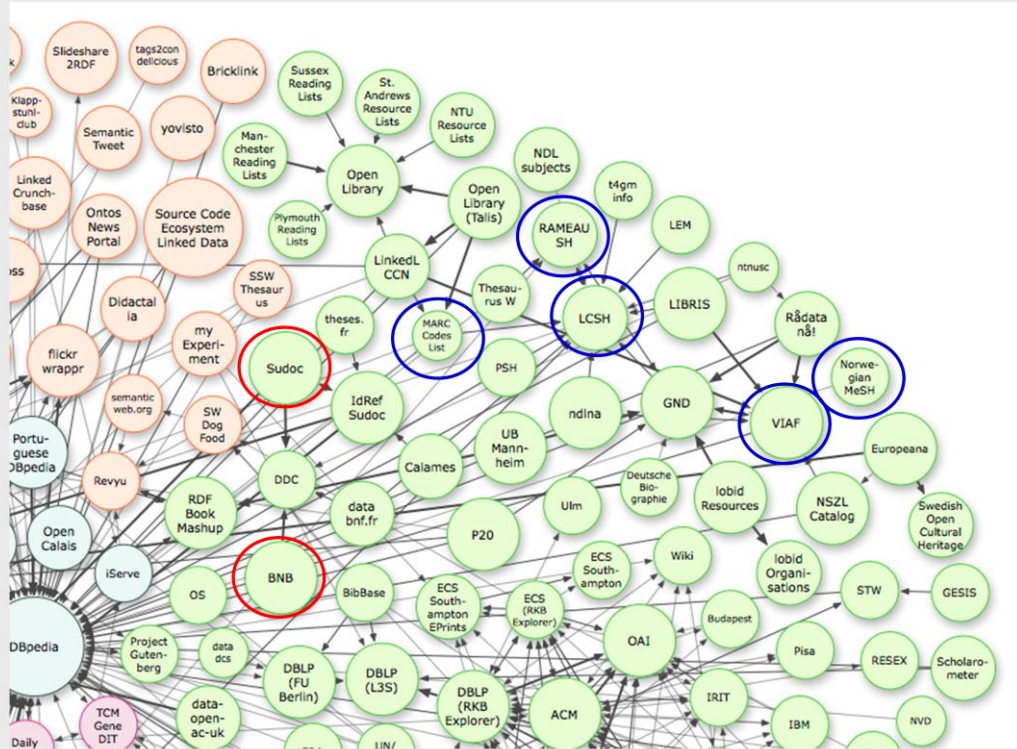
2. Французская национальная библиография

3. OCLC WorldCat

Август 2012 – OCLC объявил публикации около 1.2 млн ресурсов из каталога WorldCat (около 80 млн «троек»). Работа над публикацией WorldCat в виде связанных данных продолжается.



Отметим, что это не граф RDF: каждый узел представляет пространство имен троек RDF, а соединительные линии просто обозначают существование некой связи между пространствами имен. Облако постоянно растет.



«Библиотечный угол» LOD (пространства имен, так или иначе связанные с библиотеками, архивами, музеями, описательными и авторитетными данными):

На рисунке отмечены некоторые национальные библиографии и каталоги, а также наборы данных (авторитетные файлы и перечни кодов), уже опубликованные в виде связанных данных:

- Британская национальная библиография
- Французская национальная библиография,
- Каталог Национальной библиотеки Венгрии
- VIAF
- LCSH
- Перечни кодов MARC

<http://www.rusmarc.ru>

olga_zhlobinskaya@nlr.ru