RDF FOR LIBRARIANS

JENN RILEY
METADATA LIBRARIAN
DIGITAL LIBRARY PROGRAM

From Wikipedia...

"A collection of RDF statements intrinsically represents a labeled, directed multi-graph." http://en.wikipedia.org/wiki/Resource_Description_Framework

Learning the lingo *is* important; however for us it's probably better to start with understanding how some of the basic concepts are different than what we're used to.

That's what we're going to do today.

Structural differences

Libraries have "records"

4

```
=LDR 02578nim 2200505 a 45 0
=001 AJJ3915BM
                                                                               <?xml version="1.0" encoding="UTF-8" ?>
=008 970808p19971933enkmun\\\efi\\\\\\\\d
                                                                               <marc:collection xmlns:marc="http://www.loc.gov/MARC21/slim"</pre>
=007 sd?fmnann???ee
                                                                                   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
=024 1\$a2703192792
                                                                                   xsi:schemaLocation="http://www.loc.gov/MARC21/slim http://www.loc.gov/standards/
=028 02$aGEMM CD 9279$bPearl
                                                                                   <marc:record>
=035 \\$a(InU)AJJ3915BM
                                                                                        <marc:leader>01667mjm a2200433 a 45 0</marc:leader>
=033 10$a1933----$a193503--$a193504--$a19431123$a1944----
                                                                                        <marc:controlfield tag="001">AAK6198BM</marc:controlfield>
                                                                                        <marc:controlfield tag="008">890105p19871986qw snn 1 </marc:controlfield>
=035 \\$a(0CoLC)ocm37471228
                                                                                        <marc:controlfield tag="007">sd?fsngnnmpled</marc:controlfield>
=040 \\$aCR2$cCR2
                                                                                        <marc:datafield tag="010" ind1=" " ind2=" ">
=047 \\$abt$asa$avr$asn
                                                                                            <marc:subfield code="a"> 89750052 /R</marc:subfield>
=245 10$aAaron Copland & Leonard Bernstein$h[sound recording].
                                                                                        </marc:datafield>
=246 3\$aAaron Copland and Leonard Bernstein
                                                                                        <marc:datafield tag="028" ind1="0" ind2="1">
=246 1\$iTitle on container:$aComposer as performer
                                                                                            <marc:subfield code="a">423 065-2</marc:subfield>
                                                                                            <marc:subfield code="b">Deutsche Grammophon</marc:subfield>
=260 \\$aWadhurst, E. Sussex, England :$bPearl,$cp1997.
                                                                                        </marc:datafield>
=300 \\$a1 sound disc :$bdigital, mono. ;$c4 3/4 in.
                                                                                        <marc:datafield tag="035" ind1=" " ind2=" ">
=500 \\$aThe 1st work a ballet, originally for orchestra, reduced for piano.
                                                                                            <marc:subfield code="a">(InU)AAK6198BM</marc:subfield>
=511  O\$aAaron Copland, Leonard Bernstein, pianos ; Ethel Luening, soprano (3rd work
                                                                                        </marc:datafield>
(6th work) ; Ivor Karman (2nd work), Jacques Gordon (5th work), violins ; David Freed
                                                                                        <marc:datafield tag="033" ind1="0" ind2=" ">
=518 \\$aRecorded ca. 1933 (3rd work), in Mar., 1935 (2nd, 4th works), Apr., 1935 (5
                                                                                            <marc:subfield code="a">198607--</marc:subfield>
works), and in 1944 (1st work).
                                                                                            <marc:subfield code="b">6299</marc:subfield>
=505  O\$aAppalachian spring (24:25) ; Trio Vitebsk : study on a Jewish theme (11:27)
                                                                                            <marc:subfield code="c">C6</marc:subfield>
variations (10:48); Two pieces for violin and piano (8:06) / Copland -- Sonata for o
                                                                                        </marc:datafield>
                                                                                        <marc:datafield tag="035" ind1=" " ind2=" ">
Seven anniversaries. For my sister Shirley (:54) ; In memoriam Natalie Koussevitzky (
                                                                                            <marc:subfield code="a">(OCoLC)ocm19777353</marc:subfield>
32) / Bernstein.
                                                                                        </marc:datafield>
=650 \0$aBallets.
                                                                                        <marc:datafield tag="040" ind1=" " ind2=" ">
=650 \0$aPiano music. Arranaed.
                                                                                            <marc:subfield_code="a">DLC</marc:subfield>
=650 \0$aPiano trios.
                                                                                            <marc:subfield code="c">DLC</marc:subfield>
=650 \0$aVocalises (High voice) with piano.
                                                                                            <marc:subfield code="d">OCL</marc:subfield>
=650 \0$aVariations (Piano)
                                                                                            <marc:subfield code="d">IUL</marc:subfield>
                                                                                        </marc:datafield>
=650 \0$aViolin and piano music.
                                                                                        <marc:datafield tag="041" ind1="0" ind2=" ">
=650 \0$aSonatas (Clarinet and piano)
                                                                                            <marc:subfield code="g">gerengfreita</marc:subfield>
=650 \0$aPiano music.
                                                                                        </marc:datafield>
=700 1\$aCopland, Aaron,$d1900–1990.$4prf
                                                                                        <marc:datafield tag="045" ind1="1" ind2=" ">
=700 1\$aBernstein, Leonard,$d1918–1990$4prf
                                                                                            <marc:subfield code="b">d1877</marc:subfield>
=700 12$aCopland, Aaron,$d1900–1990.$tAppalachian spring.$pSuite;$oarr.
                                                                                            <marc:subfield code="b">d1916</marc:subfield>
=700 12$aCopland, Aaron,$d1900–1990.$tVitebsk
                                                                                        </marc:datafield>
                                                                                        <marc:datafield tag="048" ind1=" " ind2=" ">
=700 12$aCopland, Aaron,$d1900-1990.$tVocalise
                                                                                            <marc:subfield code="a">sa01</marc:subfield>
=700 12$aCopland. Aaron.$d1900–1990.$tVariations.$mpiano
                                                                                            <marc:subfield code="a">ka01</marc:subfield>
=700 12$aCopland. Aaron.$d1900—1990.$tPieces.$mviolin. piano
                                                                                        </marc:datafield>
=700 12$aBernstein, Leonard,$d1918-1990.$tSonatas,$mclarinet, piano.
                                                                                        <marc:datafield tag="050" ind1="1" ind2=" ">
=700 12$aBernstein, Leonard,$d1918-1990.$tAnniversaries$n(1943).$kSelections.
                                                                                            <marc:subfield code="a">M219</marc:subfield>
=856 41$uhttp://purl.dlib.indiana.edu/iudl/variations/sound/AJJ3915
                                                                                        </marc:datafield>
=998 \\$a8/8/1997$b9/9/1999$c9/9/1999
                                                                                        <marc:datafield tag="050" ind1="0" ind2=" ">
=596 \\$a20
                                                                                            <marc:subfield code="a">Deutsche Grammophon 423 065-2</marc:subfield>
=926 \\$aB-MUSIC$bCD$cCD ZC.66$dAV$f1
```

5

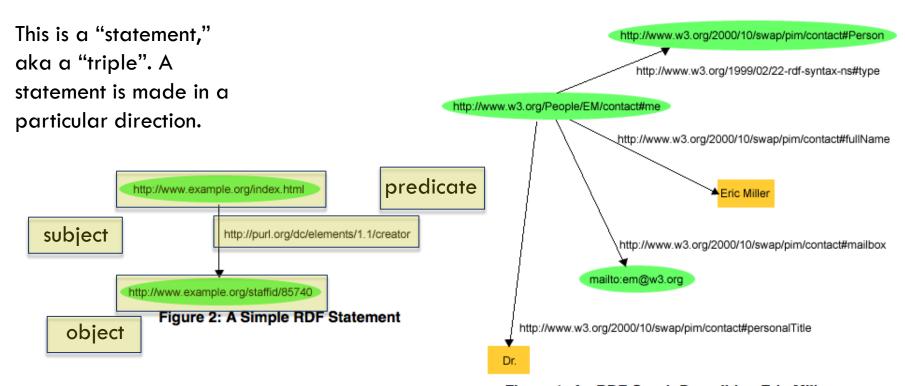


Figure 1: An RDF Graph Describing Eric Miller

Statements combine to form graphs. A graph is of no fixed size and contains no predetermined types of statements.

(The graph is the real RDF model;

a triple is a secondary representation.)

Figures from RDF Primer http://www.w3.org/TR/rdf-primer/

More about statements

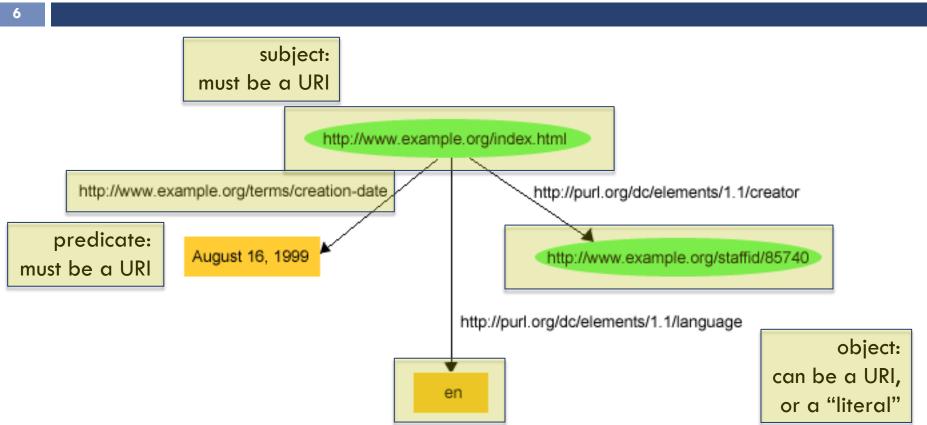


Figure 3: Several Statements About the Same Resource

7

RDF uses URIs to identify:

individuals, e.g., Eric Miller, identified by http://www.w3.org/People/EM/contact#me

kinds of things, e.g., Person, identified by http://www.w3.org/2000/10/swap/pim/contact#P erson

properties of those things, e.g., mailbox, identified by http://www.w3.org/2000/10/swap/pim/contact# mailbox

values of those properties, e.g. mailto:em@w3.org as the value of the mailbox property (RDF also uses character strings such as "Eric Miller", and values from other datatypes such as integers and dates, as the values of properties)

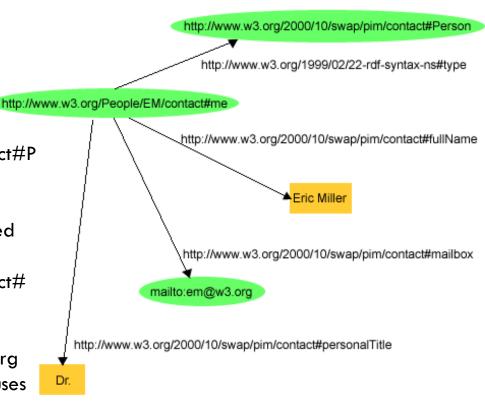


Figure 1: An RDF Graph Describing Eric Miller

Figure and text from RDF Primer http://www.w3.org/TR/rdf-primer/

Implications of statement structure (1)

- Subjects are always formalized; know exactly what is being talked about
 - They're only implicit in library metadata
 - Makes moot the 1:1 problem
 - Might still have content vs. carrier problem
- Predicates are always formalized
 - Maybe not all that different than library/archive/museum metadata models
 - More obviously reusable

Implications of statement structure (2)

- □ Representation of objects is flexible
 - Using a URI facilitates connecting this value into a larger graph
 - Literals allow more limited functionality; used when the value isn't likely to be useful as part of a vocabulary
 - URI vs. literal a design choice; note the previous example used a literal where a URI might be better!
- Making a statement that two different URIs refer to the same thing can join two graphs
- □ Note, URIs aren't necessarily dereferencable

Blank nodes

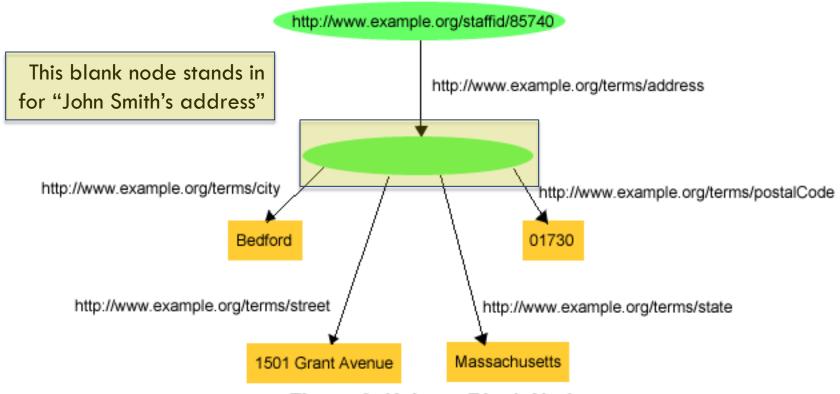


Figure 6: Using a Blank Node

Figure from RDF Primer http://www.w3.org/TR/rdf-primer/

DLP Brown Bag Series

9/22/10

Properties vs. elements

- Remember, a statement's predicate represents a property of the resource (subject) being described
 - Resource=subject; Property=predicate. I don't know why there are different terms for the model than for individual statements. Let's move on.
- On the surface, an RDF property looks like the same thing as an XML element or a database field
- But the underlying formal model is different
 - Explicit subjects and directionality of statements
 - Formality of RDF model places additional restrictions but allows more explicit meaning
- Therefore inferences you can draw for elements and fields are not as strong as you can for properties

Graph vs. tree

XML documents are trees

```
This is as good as you can
<mods version="3.4">
  <titleInfo>
                                                                                do inferring a graph from
     <title>Ekkoes from Kentucky</title>
  </titleInfo>
                                                                                           a tree:
  <name type="personal">
     <namePart>Locke, David Ross</namePart>
                                                                  XML tree model
                                                                                        RDF graph model
     <namePart type="date">1833-1888</namePart>
     <role>
                                                                       Root
         <roleTerm authority="marcrelator" type="text">creator
                                                                                                    Blank nodes
         <roleTerm type="code">cre</roleTerm>
                                                                                                     rdf:Properties
      </role>
                                                                    elém i
                                                                          elèm
  </name>
  <name type="personal">
                                                                       elem elem
                                                                elem
     <namePart>Nast, Thomas</namePart>
                                                                                attr
     <namePart type="date">1840-1902</namePart>
     <role>
                                                                       Text Empty Text
                                                               Empty.
                                                                                                Text
         <roleTerm type="text">illus.</roleTerm>
         <roleTerm type="code">ill</roleTerm>
     </role>
  </name>
</mods>
```

Model vs. its syntax(es)

- □ There's a difference between:
 - Model of information representation
 - Property or element definitions
 - Binding of the information into a specific syntax
- MARC is both the record structure (syntax) and content designations (element definitions)
- RDF model has many encoding syntaxes
- RDF and MARC operate at different levels in this landscape; but that doesn't mean MARC's structure is capable of representing the RDF model

Some realities of RDF that scare us

- No predetermined set of properties to care about
- No guarantee that the same person/item/place/whatever are always referred to with the same URI
- No inherent mechanism/requirement for vetting properties, URIs, etc
- But let's be frank here. Are our library/archive/museum records really:
 - Complete?
 - Authoritative?
 - Consistent?
 - Accurate?
 - All that functional for what we want to do?

More RDF concepts

- Class
 - A statement can say a subject is of a certain type or category known as a "class".
 - Classes can be formally defined in RDF Schema documents
- Domain and Range
 - Specify what classes subjects and objects of statements (respectively) using a given property can be
 - More than one domain or range statement can be made for any given property
- These features allow "inferencing" to deduce statements that aren't actually made

But both have...

- Raging debate over how precise you have to be in indicating what it is you're describing
- An interest in mechanisms to allow consumers of data to evaluate the authority of statements
- A tendency to over-model (see "The Modeller" blog post on handout)
- Other similarities
 - "Application profile" notion applicable to both XML and RDF models, though would be implemented differently
 - XML elements repeatable; RDF makes no restriction on how many statements about the same resource use the same predicate.

17 Terminology differences

"Subject"

- □ In libraries, what an information resource is about
- □ In RDF, what a statement is about

"Vocabulary"

- In libraries, implies a controlled vocabulary of a certain sort
 - Authorized terms
 - Lead-in terms (see references, etc.)
 - Sometimes, hierarchical structure
 - Sometimes, related terms
- □ In RDF, much looser definition
 - Includes formal definitions of classes and properties

"Class"

- In libraries, a classification scheme indicating the general topic or area of knowledge covered by an information resource
- In RDF, a type or category that any type of object or resource belongs to

"Schema"

- XML Schema defines a set of elements intended to be used together
- RDF Schema defines classes and properties intended to be used anywhere, alone or in combination

So, wait, why should I learn this?

Libraries are moving in this direction

- At least a little bit
- Increasing viability and acceptance of interoperating with data from outside of libraries
- RDA gives us an opportunity to fundamentally rethink some features of our data
- Semantic Web activities have been given new life with the grassroots Linked Data movement

Linked Data design issues

- □ Use URIs as names for things
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- Include links to other URIs. so that they can discover more things.

Tim Berners-Lee http://www.w3.org/DesignIssues/LinkedData.html

How much planning do we need to do?

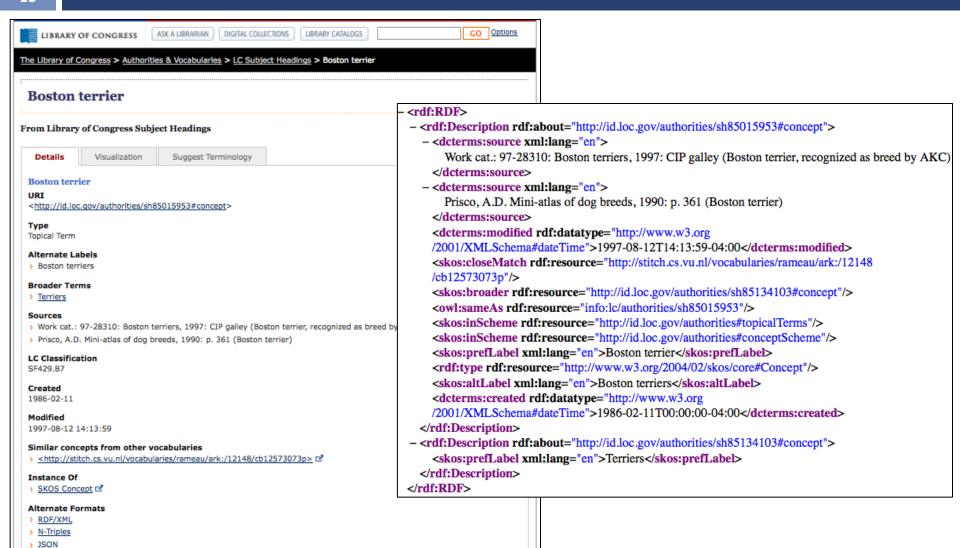
"There are other cases where the easiest thing for somebody to do is to just put data up in whatever form it's available."

```
Tim Berners-Lee <a href="http://www.readwriteweb.com/archives/interview_with_tim_b">http://www.readwriteweb.com/archives/interview_with_tim_b</a> erners-lee_part_1.php>
```

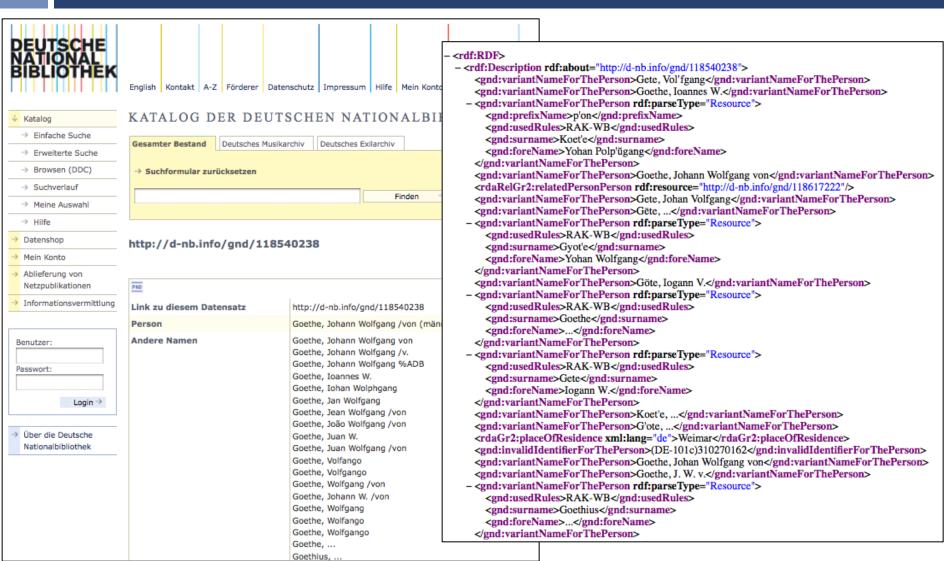
Let's review some of what libraries have done so far.

id.loc.gov

26

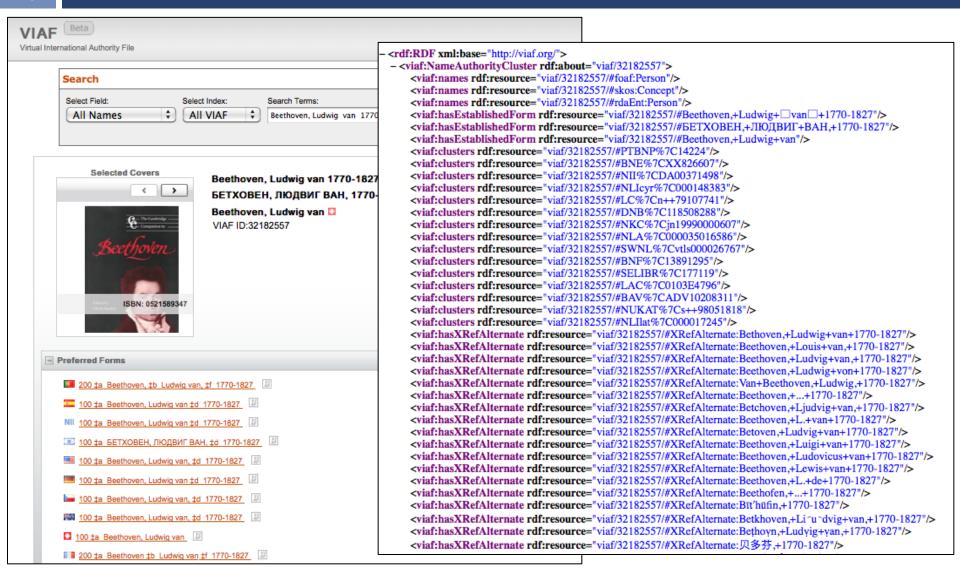


German National Library



Virtual International Authority File

28



LIBRIS – Swedish National Library

```
LIBRIS @prefix dc: <a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/> .</a>
           @prefix owl: <http://www.w3.org/2002/07/owl#> .
           @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
           @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
           @prefix libris: <a href="http://libris.kb.se/vocabulary/experimental#">http://libris.kb.se/vocabulary/experimental#</a>.
onr:5060570
           @prefix bibo: <a href="http://purl.org/ontology/bibo/">http://purl.org/ontology/bibo/> .
           <http://libris.kb.se/resource/bib/5060570>
                                                               rdfs:isDefinedBy
                                                                                      <a href="http://libris.kb.se/data/bib/5060570">http://libris.kb.se/data/bib/5060570</a>.
Search: onr:5060
           <http://libris.kb.se/resource/bib/5060570>
                                                               rdf:type
                                                                             bibo:Book .
  1 of 1 4 Pre
           <http://libris.kb.se/resource/bib/5060570>
                                                                             "The difference engine"@en .
                                                               dc:title
           <http://libris.kb.se/resource/bib/5060570>
                                                                               "Gibson, William, 1948-" .
                                                               dc:creator
                                                                               "William Gibson" .
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:creator
  The differe
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:creator
                                                                               "Sterling, Bruce, 1954-" .
           <http://libris.kb.se/resource/bib/5060570>
                                                                               "Bruce Sterling" .
  Gibson, Willia
                                                               dc:creator
  Sterling, Bruc
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:type
                                                                            "text" .
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:publisher
                                                                                  "Vista" .
  London : Vista
  English 383 s.
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:date
                                                                            "1996" .
  Book Book
                                                                                    "Originally published: London: Gollancz, 1990"@en .
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:description
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:subject
                                                                               "American fiction" .
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:subject
                                                                               "Steampunk" .
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:identifier

√URN:ISBN:0575600292> .

  Abstract
           <http://libris.kb.se/resource/bib/5060570>
                                                               bibo:isbn10
                                                                                "0575600292" .
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:creator
                                                                               <http://libris.kb.se/resource/auth/220040> .
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:creator
                                                                               <http://libris.kb.se/resource/auth/307779> .
   Get it
           <http://libris.kb.se/resource/bib/5060570>
                                                               dc:subject
                                                                               <http://libris.kb.se/resource/auth/308073> .
  Loan Interlif
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:subject
                                                                               <http://libris.kb.se/resource/auth/308073#concept> .
           <http://libris.kb.se/resource/bib/5060570>
                                                               do:subject
                                                                               <http://libris.kb.se/resource/auth/308074> .
  Title available
           <http://libris.kb.se/resource/bib/5060570>
                                                                               <http://libris.kb.se/resource/auth/308074#concept> .
                                                               dc:subject
 ▼ Southern Swe
           <http://libris.kb.se/resource/bib/5060570>
                                                                                    <http://libris.kb.se/resource/library/Ld> .
                                                               libris:held_by
  Folkbiblioteke
           <http://libris.kb.se/resource/bib/5060570>
                                                               libris:held_by
                                                                                    <http://libris.kb.se/resource/library/Li> .
  O Local librar

¬http://libris.kb.se/resource/bib/5060570> libris:frbr_related ¬http://libris.kb.se/resource/bib/5059476> .

  Loan status . Opening hours, address etc.
```

Open Library





5 editions of Here comes everybody by Clay Shirky Add another?

← Previous | Next →



Manage Covers

Here Comes Everybody The Power of Organizing With

Published **February 28, 2008** by <u>Penguin</u> Written in <u>English</u>.

The Physical Object

| Format | Hardcover |
|-----------------|-----------|
| Number of pages | 336 |

ID Numbers

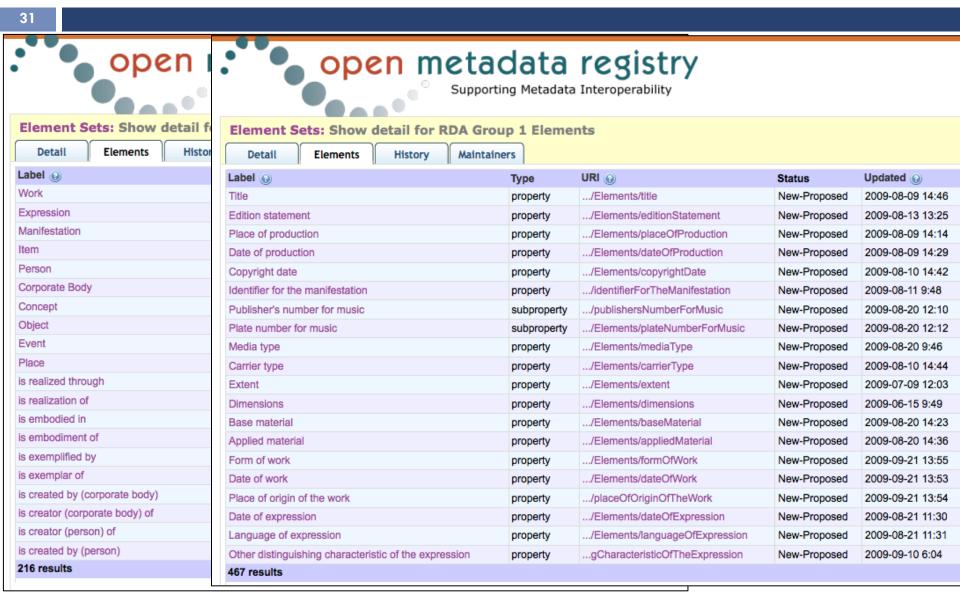
| Open Library | OL12433648M |
|---------------|---------------|
| ISBN 10 | 1594201536 |
| ISBN 13 | 9781594201530 |
| Library Thing | 4312391 |
| Goodreads | 1998185 |
| | g Add book |

2,248 people read it (3.3

```
<rdf:RDF>
- <rdf:Description rdf:about="http://openlibrary.org/books/OL12433648M">
    <!-- authors -->
  - <bibo:authorList rdf:parseType="Collection">
    - <rdf:Description rdf:about="http://openlibrary.org/authors/OL482102A">
        <rdf:value>Clay Shirky</rdf:value>
      </rdf:Description>
    </bibo:authorList>
    <!-- bibliographic description -->
  - <dcterms:title>
      Here Comes Everybody The Power of Organizing Without Organizations
    </dcterms:title>
    <dcterms:publisher>Penguin Press HC</dcterms:publisher>
    <dcterms:issued>February 28, 2008</dcterms:issued>
    <dcterms:extent>336</dcterms:extent>
    <!-- subjects -->
    <dc:subject>Business & Economics</dc:subject>
    <dc:subject>Business / Economics / Finance</dc:subject>
    <dc:subject>Business/Economics</dc:subject>
    <dc:subject>Management - General</dc:subject>
    <dc:subject>Telecommunications</dc:subject>
    <dc:subject>Business & Economics / Management</dc:subject>
    <dc:subject>Computer networks</dc:subject>
    <dc:subject>Information Technology</dc:subject>
    <dc:subject>Internet</dc:subject>
    <dc:subject>Social aspects</dc:subject>
    <!-- description -->
    <dcterms:language>eng</dcterms:language>
    <!-- identifiers -->
    <dcterms:identifier>/books/OL12433648M</dcterms:identifier>
    <bibo:isbn10>1594201536</bibo:isbn10>
    <bibo:isbn13>9781594201530</bibo:isbn13>
    <!-- administrative -->
    <dcterms:modified>2010-08-17 12:56:29.759962</dcterms:modified>
  </rdf:Description>
</rdf:RDF>
```

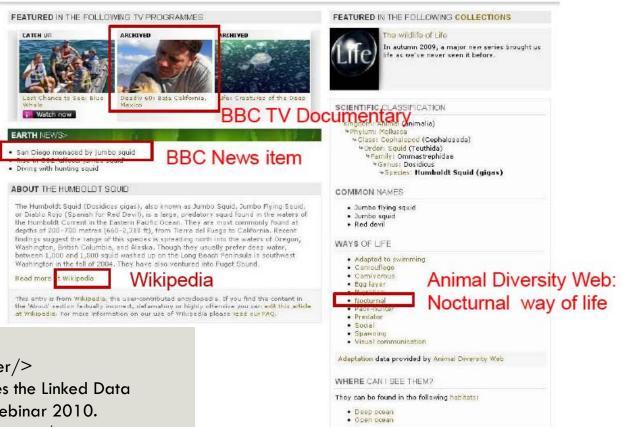
Registering FRBR and RDA properties

Registering i RDR and RDA properties





Humboldt Squid page, pulled together from a diversity of Linked Data sources



ELSEWHERE ON THE WEB

a Colonal Diversity Web Lanimaldiversity upper uppirb adult

BBC Wildlife Finder

http://www.bbc.co.uk/wildlifefinder/

Slide by Thomas Baker, "What Makes the Linked Data

Approach Different," NISO DCMI Webinar 2010.

http://dublincore.org/resources/training/NISO_Webinar_

20100825/dcmi-webinar-02.pdf>

Challenges to implementing RDF in practice

How do you...

- □ find all the triples you need?
- know what predicates and URIs to use when creating new triples?
- know what predicates and URIs to use when processing data?

What infrastructure needs to be built in order to...

- □ negate an erroneous statement?
- say that a statement is time-dependent?
- judge the likely validity of a statement made by someone else?

...actually end up with machine-understandable data in the end???

Don't run away!

- These uncertainties do not give librarians permission to write off the entire model.
- Incorporating new ideas doesn't mean we have to give up our core principles.
- □ We have an opportunity here:
 - to create classes and properties that represent data as we think it should be.
 - to bring new features to the model and to its implementations.
 - to provide some of the infrastructure that's missing.
 - to participate in the discussion.

Let's go!



For more information

- □ jenlrile@indiana.edu
- These presentation slides: http://www.dlib.indiana.edu/~jenlrile/presentations/bbfall10/rdf /rdfForLibrarians.pptx>
- Recording of talk on DLP Brown Bag page: http://www.dlib.indiana.edu/education/brownbags/>
- More resources on handout: http://www.dlib.indiana.edu/~jenlrile/presentations/bbfall10/rdf/handout.pdf

Thank you!