

Library Data and Its Future on the Web

Sandy Rodriguez

University of Missouri—Kansas City

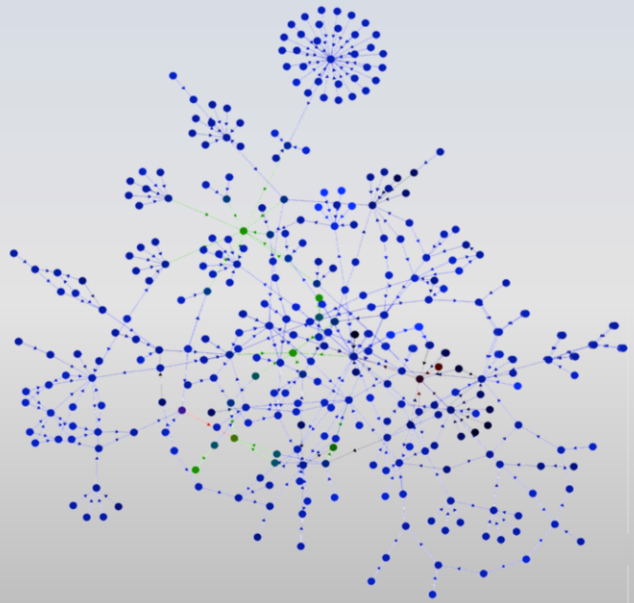
Felicity Dykas

University of Missouri—Columbia

Mary Aycock

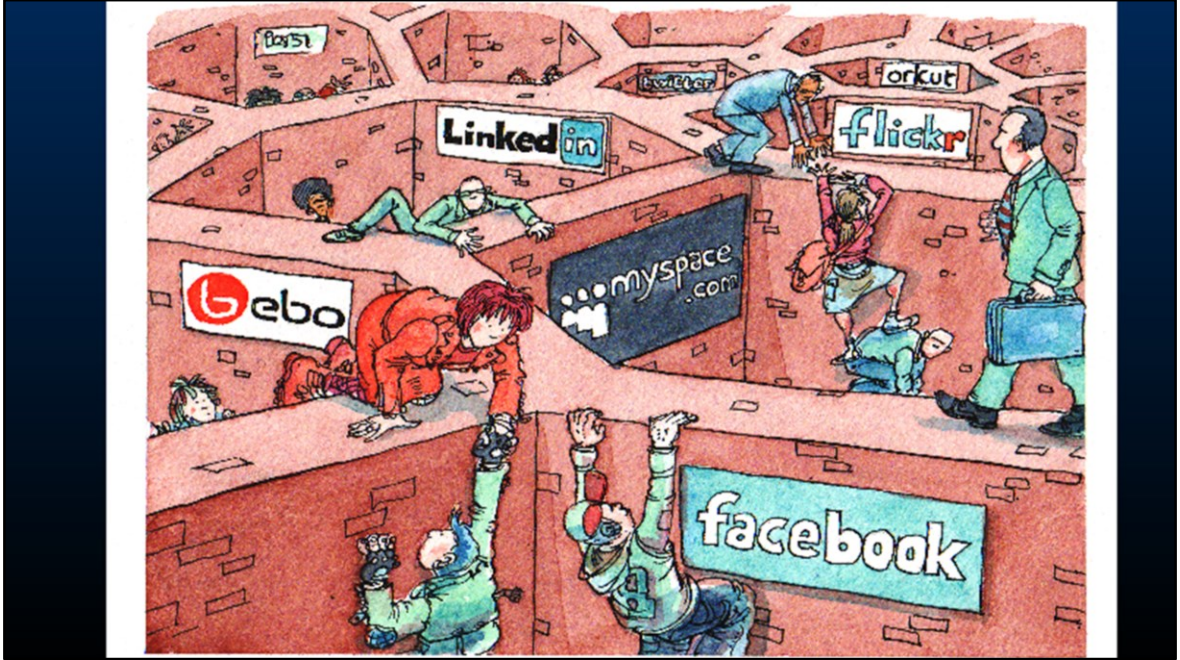
Missouri University of Science & Technology

MOBIUS Annual Conference 2015



Outline

- Linked Data and Linked Open Data
- BIBFRAME
- Linked Data in Libraries



But we've got this problem: we silo off our good quality data and struggle to expose that data in the *place* that a large population of our users are – or at least where they start. There's a bit of a language and packaging problem to put it simplistically, and we're not the only ones looking to solve this problem. This image is from Tim Berners-Lee's TED talk on The Next Web (link provided in reference slide) in which he references social media sites' need to share data across platforms. But keep in mind that this was six years ago, and much has changed since then.

So there's this movement afoot – it's here already and you've seen it already.

Image from <http://www.w3.org/2009/Talks/0422-www2009-tbl/#%281%29>.

The image shows a Google search interface for the query "convert feet to meters". The search results include a unit converter showing "1 Foot" equals "0.3048 Meter". Below the converter, there are two "Enriched Data" snippets for "Green Papaya Salad" from ThaiTable.com and norecipes.com. To the right, a "Disambiguation" box shows results for "Bonobo (Musician)" and "Bonobo (Primate)".

Quick Answers (indicated by an orange arrow pointing to the unit converter)

Enriched Data (indicated by an orange arrow pointing to the recipe snippets)


Disambiguation (indicated by an orange arrow pointing to the Bonobo results)

For instance, Google is using linked data to provide services like quick answers, disambiguation, rich snippets....

Tim Berners-Lee - World Wide Web Consortium
www.w3.org/People/Berners-Lee/ ▾ World Wide Web Consortium ▾
 Tim Berners-Lee invented the World Wide Web in 1989. He founded and Directs the World Wide Consortium (W3C) the forum for technical development of the ...
[Longer Biography](#) - [Answers for young people](#) - [Frequently Asked Questions](#) - [Card](#)

Tim Berners-Lee - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Tim_Berners-Lee ▾ Wikipedia ▾
 Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FEng, FRSA, DFBCS (born 8 June 1955), also known as TimBL, is an English computer scientist, best ...
[Awards and honours](#) ... - [DFBCS](#) - [Conway Berners-Lee](#) - [Mary Lee Woods](#)


In the news

 **Tech Time Warp of the Week: A 1930s Futurist Envisioned the Web With a 'Televised Book'**
 Wired - 2 days ago
 Tim Berners Lee created the world wide web in 1989. But the idea of hypertext stretches ...

[More news for tim berners lee](#)

Tim Berners-Lee | Internet Hall of Fame
internethalloffame.org ▾ Inductees ▾ Internet Hall of Fame ▾
 In 1989, Tim Berners-Lee invented the World Wide Web, an Internet-based hypermedia initiative for global information sharing while at CERN, the European ...

Tim Berners-Lee | Speaker | TED.com
<https://www.ted.com/speakers/tim-berners-lee> ▾ TED ▾



Tim Berners-Lee
 Computer Scientist

Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FEng, FRSA, DFBCS, also known as TimBL, is an English computer scientist, best known as the inventor of the World Wide Web. [Wikipedia](#)

Born: June 8, 1955 (age 59), London, United Kingdom
Spouse: Rosemary Leith (m. 2014)
Organizations founded: World Wide Web Consortium, World Wide Web Foundation
Books: *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*
Parents: Mary Lee Woods, Conway Berners-Lee
Awards: Marconi Prize, MacArthur Fellowship, Charles Stark Draper Prize, Mountbatten Medal, President's Medal

Profiles

And is even returning “things” like these knowledge cards created on the fly. It’s not just one piece of data, but a whole set of data returned, images, birth dates, organizations founded, etc.

So we absolutely have a place in the linked data movement given our legacy of rich quality data creation. There’s a lot of momentum in terms of library data. Felicity will be talking specifically unlocking our data with BIBFRAME and Mary will cover some library linked data projects underway.

But first...

Linked Open Data: What Is It?



We should step back and take a moment to understand what linked data is and why it's important to cultural memory organizations. [Watch the video.]

Don't
forget
the
bots



Simple enough, right?

Linked data leverages the web infrastructure to create a global knowledge repository allowing links and data sharing that both humans *and machines* can consume.

Designing for the bot will serve to our benefit because this kind of design where defined things are linked to other defined things through defined relationships gives us greater power. Using this technology enables us to build tools, facilitate further discovery and understanding, the possibilities are endless...

LINKED DATA

- ★ On the web, open license
 - ★ ★ Machine-readable data
 - ★ ★ ★ Non-proprietary format
 - ★ ★ ★ ★ RDF standards
 - ★ ★ ★ ★ ★ Linked RDF
- IS YOUR DATA 5 ★ ?



A good basic guide to creating linked data is the 5 star open data deployment scheme. You've probably seen this before in different iterations. The mug brings out the importance of that data being OPEN. So the first three are pretty self-explanatory and should be familiar to everyone in this room. It's when you get to RDF standards and linked RDF that you might have some questions so I want to cover that briefly because it's important to have a basic understanding of RDF.



= Resource Description Framework

Standard model for publishing and linking data on the web



Designed for:



RDF stands for the Resource Description Framework and is the standard model for publishing and linking data on the web. That is, for expressing information about resources. And resources can be anything--documents, people, physical objects, and abstract concepts. I want to emphasize that it's used for situations in which information on the Web needs to be processed **by applications**. **Not for human consumption**.

So please don't worry when you see the next several slides. You won't actually have to look at or create this complicated serialized data; instead user-friendly tools will be built around these standards. If I've got time, I hope to show you a few examples of tools that leverage linked data. However, it's still beneficial to gain a basic understanding of **how** it all works.

RDF Triples



RDF is based on the idea of making statements about resources. These statements are known as “triples.”

RDF Triples

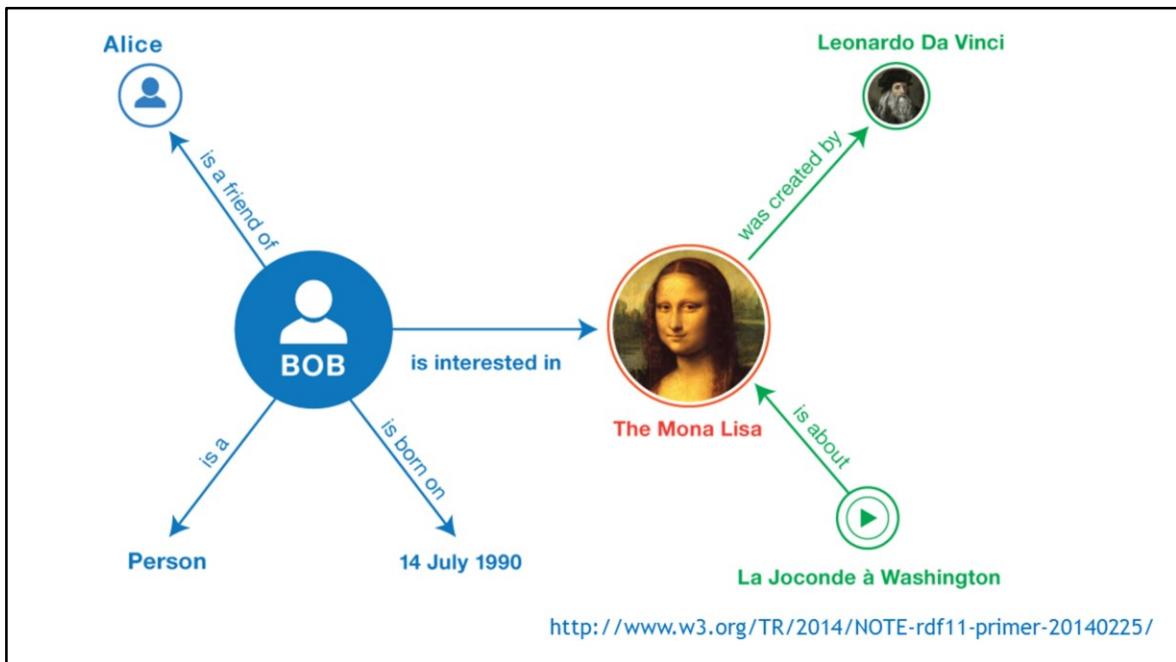
three components:

SUBJECT

PREDICATE

OBJECT

Some thing has some relationship to some other thing.



Here's an example:

- Bob is a person
- Bob is a friend of Alice
- Bob was born on the fourth of July 1990
- Bob is interested in the Mona Lisa
- The Mona Lisa was created by Leonardo da Vinci
- The video 'La Joconde à Washington' is about the Mona Lisa

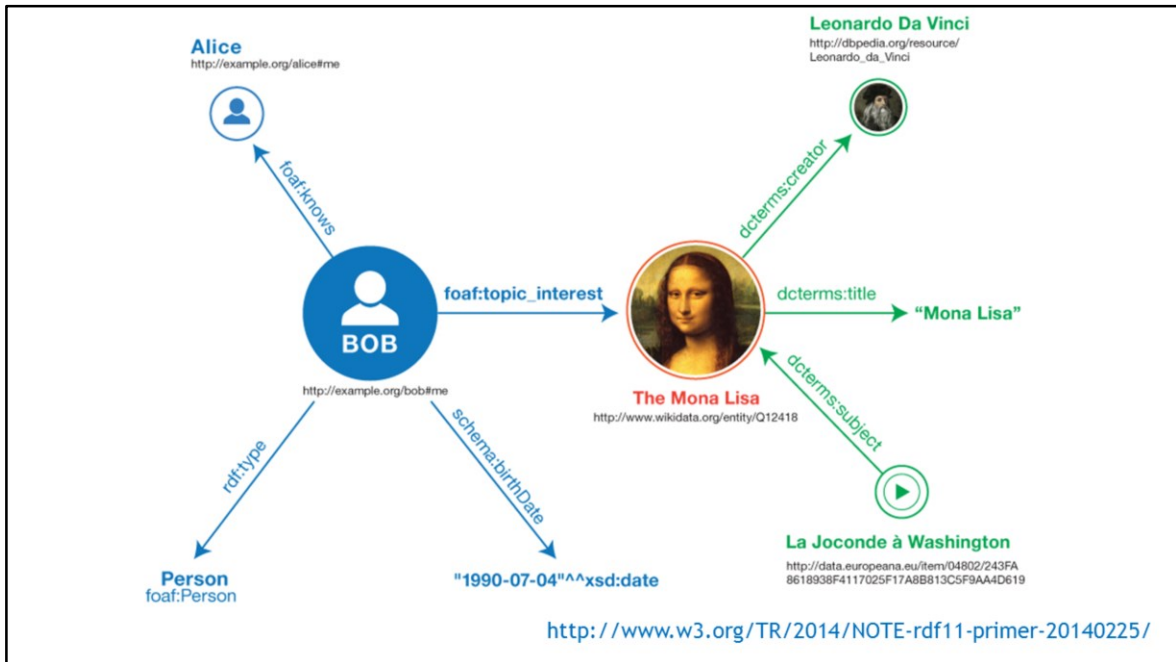
This is looking familiar, right? Due to the nature of triples and linkages, you can start to see that we are building a graph. Graph data presents a nice visualization of how these statements relate to each other, before looking at how they might be expressed in RDF. But what we have here are natural language statements of the subject, predicates, and objects.

Principles of Linked Data

1. Use URIs to name (identify) things.
2. Use HTTP URIs so that these things can be referred to and looked up (“dereferenced”) by people and systems.
3. Provide useful information when a URI is dereferenced, using open standards such as RDF.
4. Include links to other related things, using their URIs, when publishing data to the Web.

<http://www.w3.org/DesignIssues/LinkedData.html>

And if we look back at the principles of linked data outlined nicely in the short video, we need to do more. URIs are, of course, Uniform Resource Identifiers. [Read the slide.]



So looking back at that same graph data, we now see that we have URIs for the subjects and objects (Alice, Bob, Da Vinci's dbpedia page – converts structured information from Wikipedia's info boxes into linked data), and we also have literals such as `dcterms:title = "Mona Lisa"` which is allowed for Objects.

We also have a variety of selected ontologies in use for the predicates – defining some relationship (FOAF, Schema, dcterms). I should note that while it looks like the predicates here are not HTTP URIs, they are. The selected schema followed by a colon and then the vocabulary term is shorthand because the HTTP URI is specified elsewhere so that you may just use the short prefix (e.g., `foaf: dcterms:`).

RDF is a model notation

Common Serializations:

N-triples

Terse RDF Triple Language

RDF/XML

JSON-LD

RDFa

N-triples

```
<http://www.wikidata.org/entity/Q12418>  
<http://purl.org/dc/terms/creator>  
<http://dbpedia.org/resource/Leonardo_da_Vinci> .
```

Turtle

```
@prefix dc: <http://purl.org/dc/terms/> .  
  
<http://www.wikidata.org/entity/Q12418> dc:creator  
<http://dbpedia.org/resource/Leonardo_da_Vinci> .
```

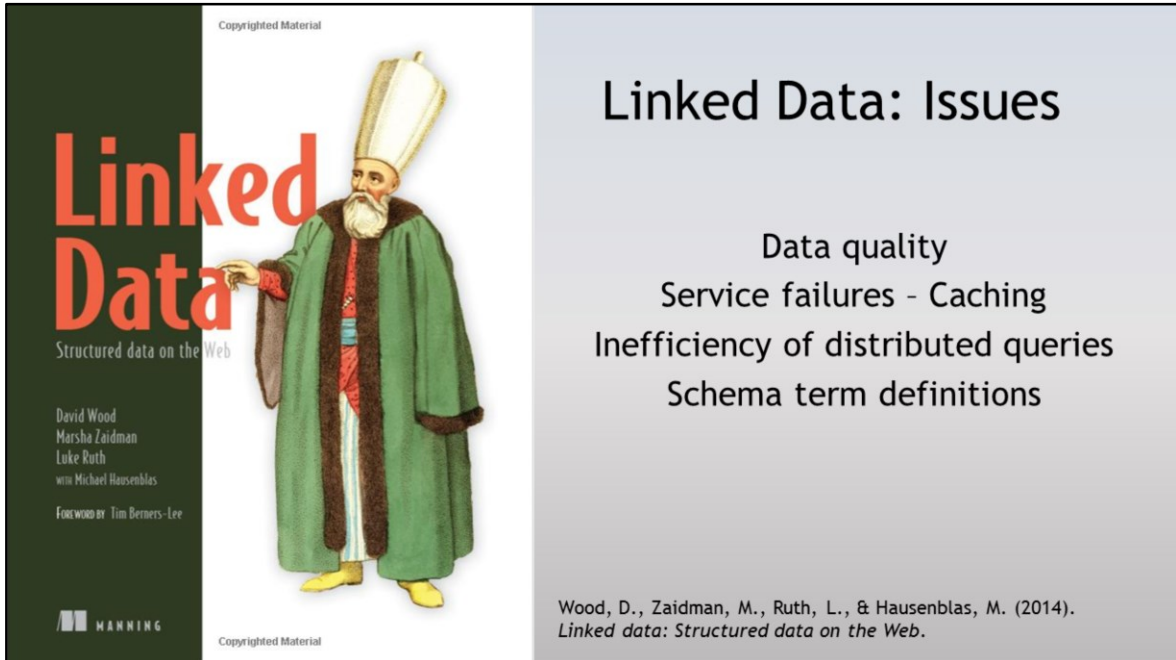
<http://rdf-translator.appspot.com/>

RDF is a model not a notation.

There are five common RDF “serializations” (i.e. ways to represent the data specified):

- N-triples (as above) - easiest to write, but can be very verbose and hard to understand
- Terse RDF Triple Language (or Turtle, is an abbreviated form of N-triples) - most human readable format
- RDF/XML (the original serialization) - can be hard to read, but good for systems that are XML-based
- JSON-LD (Javascript Object Notation for Linked Data) - common format for many modern systems; hard to read
- RDF in Attributes - RDF embedded in attributes in HTML pages

RDF triples can be expressed in **any** of these formats, and be translated from one to another. I encourage you to play around with the rdf translator if you're interested in serializations.



As the authors of this book note, “Linked Data is no silver bullet” – it inherits the same issues that the web has. It won’t protect you from bad data, from service failures, doesn’t efficiently address the issue of distributed queries; if schema term definitions change over time, it may make your data confusing or difficult to understand.

There are other issues in addition to this list, and I’ve provided a citation to a paper by MU’s Heather Moulaison and A.J. Million in our references.

Linked Data Tools



As I said, raw serialized RDF is not what you as a librarian will be constructing. There are a number of tools that have been built and that you can use in your work now that makes use of linked data, and there are tools that have been developed to be adapted for use by others.

Google refine VIAFMatching-Aug2013 [Permalink](#)

Facet / Filter Undo / Redo 127

Refresh Reset All Remove All

2057 matching records (5076 total) Extens

Show as: rows records Show: 5 10 25 50 records

ConcatName: judgment change invert reset

4 choices Sort by: name count

(unreconciled) 5

matched 2074 exclude

new 3

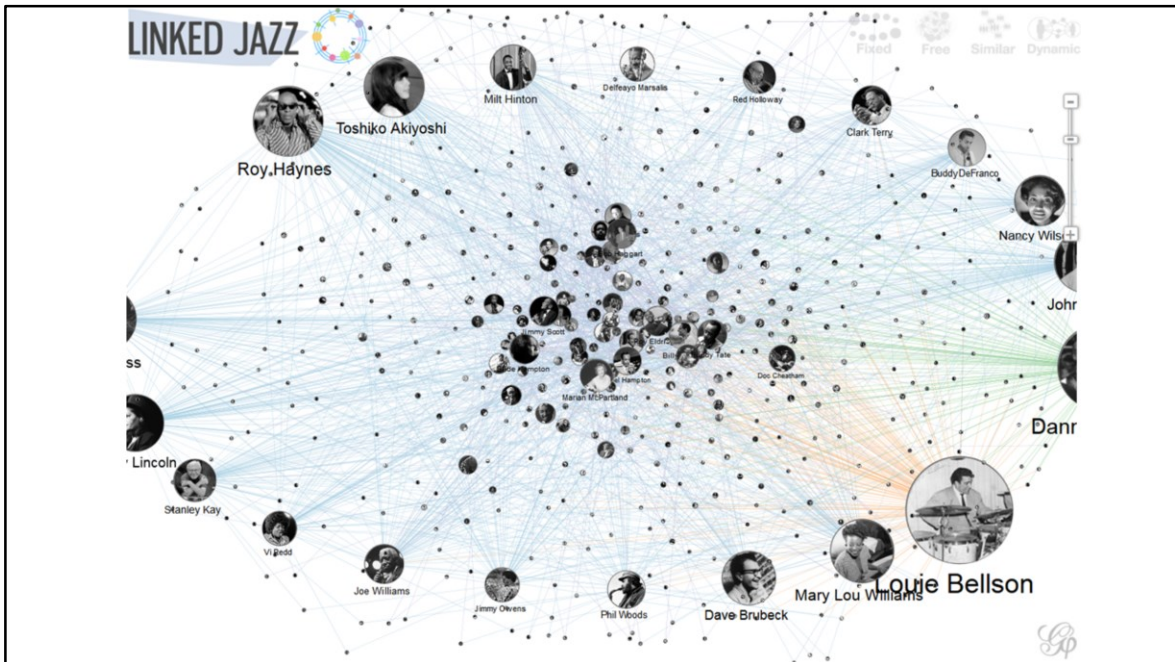
none 3416

Facet by choice counts

All	?FamilyName	?GivenName	?Dates	ConcatName	?Label
☆	343. Bell	Quentin Claudian Stephen	1910-1996	Bell, Quentin, 1910-1996 Choose new match	Bell, Quentin Claudian Stephen, 1910-1996, Author Art Critic
☆				Bell, Sir, Charles, 1774-1842, Knight, surgeon <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic	Bell, Sir, Charles, 1774-1842, Knight, surgeon
☆	349. Bell	William Blair-	1871-1936	Blair-bell, William Choose new match	Bell, William Blair-, 1871-1936, British gynaecologist and obstetrician
☆				Bellairs, Carlyon, 1871-1955, RN Commander, politician <input checked="" type="checkbox"/> <input checked="" type="checkbox"/> Create new topic	Bellairs, Carlyon, 1871-1955, RN Commander, politician edit

<http://blog.archiveshub.ac.uk/2013/08/16/hub-viaf-namematching/>

For instance, Open Refine is a powerful data cleanup and transformation tool, and it has a reconciliation service which allows you to query open data sets which means that if that data is an openly published linked data set than you can use it to reconcile things like names. This screenshot shows the querying of VIAF (Virtual International Authority File). You can see the hyperlinks indicate a match in VIAF. OpenRefine is fairly easy to learn.



The Linked Jazz project uses Linked Open Data to enhance discoverability and visibility, and show relationships between jazz musicians (<https://linkedjazz.org/network/>). They mined and extracted names from transcripts to expose these relationships – so if a jazz musician mentions another jazz musician, they used an ontology (FOAF) to expose the relationship as “knows of” - then, they crowdsource more explicit, specific relationship (mentoring) by asking folks to read the transcript.

Linked Jazz - Developing Tools

Gene Ammons

LINKED JAZZ

Perfect (965) High (1,319) Medium (2,498) Low (1,725) Many (1,256) None (1,089) | Verified (13) Deleted(1) | Search

	Martin "Chink" Abraham	1886 - 1981					
	Nathan Abshire	1913 - 1981					
	Buddy Ace	1936 - 1994					
	Pepper Adams	1930 - 1986					
	Nat Adderley	1931 - 2000					
	Joe Albany	1924 - 1988					
	Alvin Alcorn	1912 - 2003					
	Alger Texas Alexander	1900 - 1954					
	Roland Alexander	1935 - 2006					
	Rashied Ali	1933 - 2009					

Pepper Adams

Park Frederick "Pepper" Adams III (October 8, 1930 - September 10, 1986) was a jazz baritone saxophonist and composer. He composed 43 pieces, was the leader on eighteen albums spanning 28 years, and participated in 600 sessions as a sideman.

LIBRARY OF CONGRESS
<http://id.loc.gov/authorities/names/n79115718>
altLabel | Adams, Park, 1930-1986
prefLabel | Adams, Pepper, 1930-1986

So they developed many tools which leverage linked data including:

- the Name Mapping and Curator Tool which created a directory of jazz artists, maps the names to Dbpedia and then maps individuals' URIs onto the LCNAF and VIAF to include preferred and alternate names. The Curator, is a user-friendly interface for this heavily automated process, but it allows for human curation of the directory, including the approval, removal and disambiguation of personal names.
- Linked Jazz 52nd street is their crowdsourcing tool – it allows jazz experts and enthusiasts to read the transcripts more closely and more specifically identify the relationship beyond a “knows of” relationship. The identified relationship is converted into RDF statement that feed the project's LOD dataset.

*Pratt Institute with funding from OCLC and cooperation with Rutgers Jazz

**Be
Bold**



So I'll conclude by saying that you should be bold. Start small, do a little learning, experiment, and use some available tools.

Is there a place for libraries in a linked data environment?

Yes! We want our library resources to be discoverable on the web. Bibliographic information in our institutional repositories and digital libraries may be harvested (OAI) and available. **Bibliographic information in our OPACs is not.**



Studies show that our users have moved to the web as a starting point for research and for other information needs. We need to make sure they find links to our resources on the Web. A major initiative to make that happen is BIBFRAME. I will cover some basics about BIBFRAME and give you an update on its current status.

Bibliographic Framework Initiative

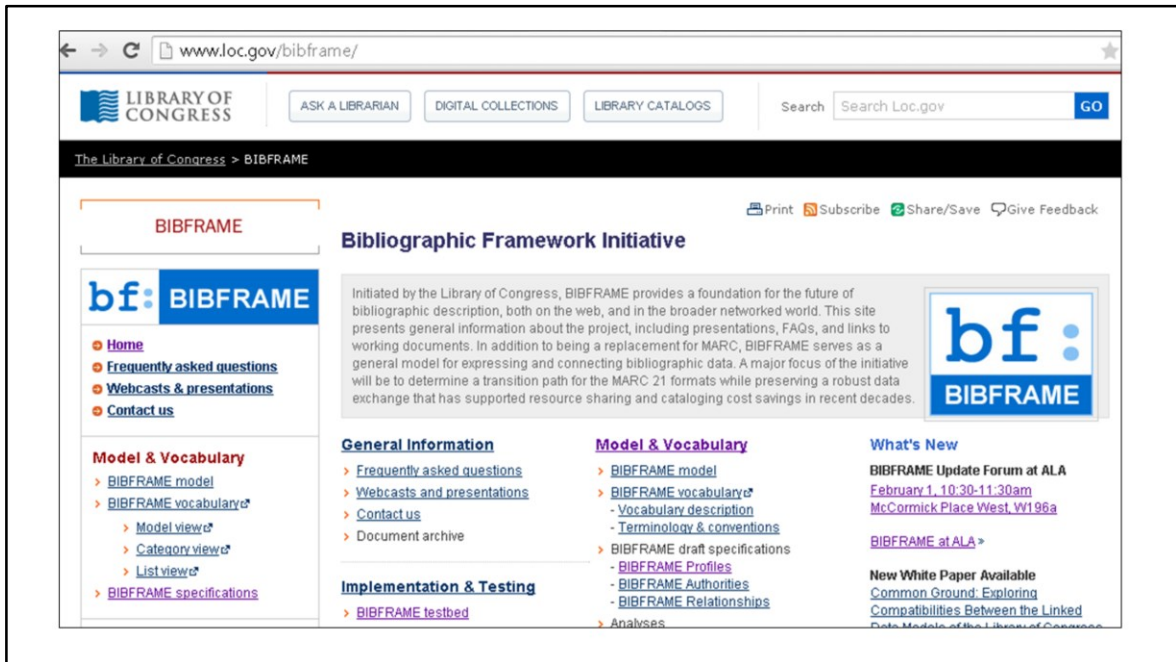
- Initiated by the Library of Congress
- Development of a bibliographic framework, called BIBFRAME
 - provide a foundation for the future of bibliographic description, both on the web, and in the broader networked world.
 - replacement for MARC
 - serve as a general model for expressing and connecting bibliographic data
 - A major focus of the initiative will be to determine a transition path for the MARC 21 formats while preserving a robust data exchange that has supported resource sharing and cataloging cost savings in recent decades.

<http://www.loc.gov/bibframe/>

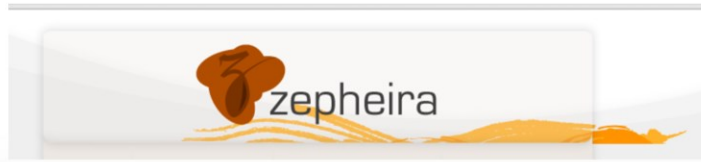
First, what is BIBFRAME? Information from the LC site. BIBFRAME is the result of an initiative by the Library of Congress. In May 2011, LC issued a statement about Transforming our Bibliographic Framework. It outlined steps the Library of Congress would take “to analyze the present and future environment, identify the components of the framework to support our users, and plan for the evolution from our present framework to the future—not just for the Library of Congress, but for all institutions that depend on bibliographic data shared by the Library and its partners.”

<http://www.loc.gov/bibframe/news/framework-051311.html>

BIBFRAME will help us to create and present data elements in our bibliographic records so that they can be put together in ways that will work on the web. This will allow users to find the resources described by them on the web.



BIBFRAME is a Library of Congress initiative. The Library of Congress website is a place where you will find background information, progress reports, details of the BIBFRAME model, and links to resources about BIBFRAME.
www.loc.gov/bibframe.



- **Zepheira is a privately held company** driven by a team of knowledge management and specific domain area experts ...
- Zepheira's expert team has led the **development and deployment of the Semantic Web** ...
- ... has provided much of the vision and technical foundation for the next generation of the Web.
- ... deeply experienced in successfully implementing these technologies ...

<https://zepheira.com/about/>

As part of its initiative, the Library of Congress issued a contract to Zepheira to develop a model. This slide shows a bit about Zepheira. LC brought in a team of experts in semantic web operations and standards. Even after their contract with LC ended, Zepheira remains heavily involved in research and experimentation.

Result of Zepheira's Work

- “Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services”
November 21, 2012
- 42 pages
- **New, proposed model called BIBFRAME**
- 18 pages about the model, including illustrations and examples
- 5 pages about linked data
- 9 pages about related library initiatives
- Conclusion
- Glossary

The result of Zepheira's work was the publication of a document which gives the outline of a system to replace MARC. It is called Bibliographic Framework, or BIBFRAME, and was issued in 2012. I give you some details of the document here so you can see that this is not a comprehensive document on BIBFRAME. The section on the BIBFRAME model, including illustrations and examples, is 18 pages long. As stated in the conclusion: ... This document outlines an initial model for the interchange of data in a Linked Data environment.

#1. BIBFRAME is not a ready-to-go, off-the shelf standard that you can pick up and start using

Library of Congress FAQ

- **“When should we move to BIBFRAME?”**

- BIBFRAME is far from an environment that you could move to yet. The model and its components are still in discussion and development -- a work in progress. When it is more mature, vendors and suppliers will need time to adjust services to accommodate it. And then we can expect a mixed environment for some time.”

<http://www.loc.gov/bibframe/faqs/#q08>

#1 important point: BIBFRAME is not a ready-to-go, off-the-shelf standard ... The 2012 document covered broad ideas that need to be studied, tested, and expanded.



“So, Can I Ignore it for now?”
No. Stay alert!

Although the **BIBFRAME** model is a draft and expected to change, the Library of Congress wants to share it now with the **community** not only so that it is **informed of progress** being made but also to engender conversation and constructive feedback. The Library is leading this initiative, but it is important that the **library community as a whole work in concert** to create an environment for bibliographic description and data exchange that recognizes and **leverages the resources and scale of a global network of data.**

Bibliographic Framework as a Web of Data, 2012, Introduction

Catalogers spent years learning about and waiting for RDA and may want to sit this one out. You may ask, Can I ignore this for now? Wake me up when it is ready to go. So, I am glad you are here. And this isn't just for catalogers. This involves IT, public services, and others. We need to work together on ways to make our bibliographic data web-friendly and user-friendly.

Following the publication of the document, LC and others began testing the ideas in BIBFRAME. In June 2014, LC announced its desire to collaborate with PCC in the endorsement and support of BIBFRAME as the model to help the library community move into the Linked Data environment. So BIBFRAME has been found to be a workable replacement for MARC.

BIBFRAME Includes ...

- BIBFRAME model
- BIBFRAME vocabularies
 - Classes
 - Properties
- BIBFRAME profiles

The following information comes from the report and other documents based on experimentation with the ideas in the report.

Work (Creative Work)	Instance	Authority	Annotation
Covers FRBR Work and Expression	FRBR Manifestation and Item	A thing or concept associated with a BIBFRAME Work or Instance	Used to give more information about a BIBFRAME Work, Instance, or Authority
The conceptual essence of a resource	An individual, material embodiment of a Work	Provides access points	Examples: Library holdings, Reviews, Cover Images, Tables of Contents, Descriptions
Serves as a Web based control point and reference point	Can be physical or digital	Includes: Agents (people, organizations, and jurisdictions), Places, Subjects	

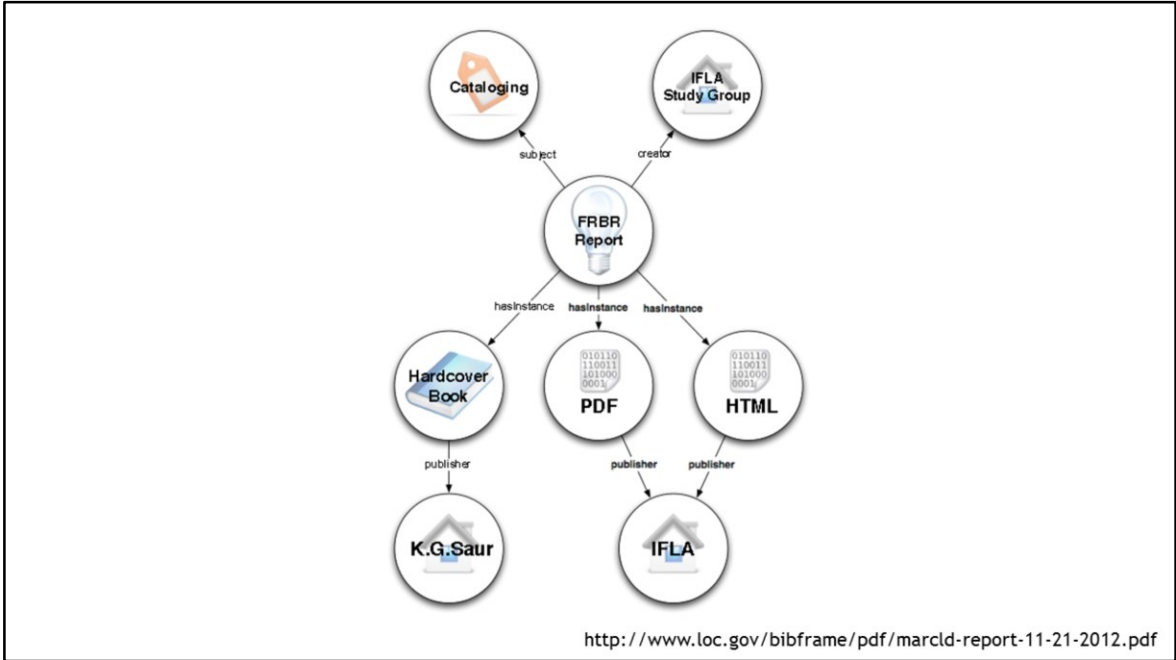
FRBR introduced four bibliographic entities. BIBFRAME collapses these into two.

- Works are the abstract idea of a bibliographic resource. The Work entity for Hamlet, for example, expresses the idea of all the various editions and translations of Hamlet. It is a reference point. Connected to that Work will be records for Instances.
- An Instance is the individual material embodiment of a Work. The 3rd English edition, published by Wilson in 1995. OR the pdf version published in 2010. An Instance can be physical or digital.
- An Authority represents subjects and those responsible for the creation of the Work or Instance. Authors, illustrators, publishers, etc.
- Annotations will add information and, in a linked data environment, this data may come from third parties. We'll get tables of contents from publishers, and reviews from journals. In the initial version of BIBFRAME, library holdings have been considered an Annotation. There is now discussion about making holdings information a separate class or category. They don't fit well with the other types of Annotations.

BIBFRAME Properties (Selected)

Work (Creative Work)	Instance	Authority	Annotation
absorbedBy classificationLcc contributor creator genre language series subject title workTitle	accompaniedBy colorContent contentNotes duration edition extent formatOfMusic frequency isbn10 title	Authorized AccessPoint identifier label relatedTo	annotates annotation AssertedBy

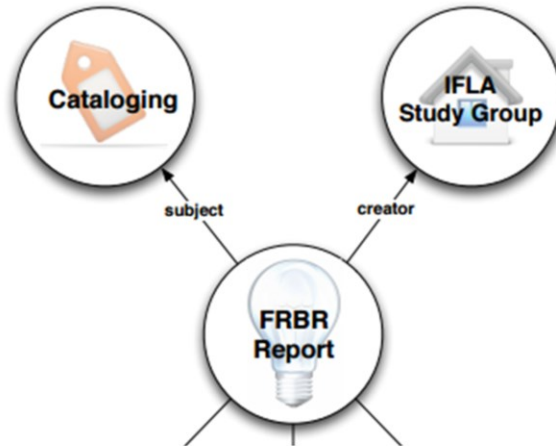
Each class defined in BIBFRAME has properties. Names are generally short and descriptive, not abstract. You don't need a translator to interpret these, as you might need with MARC. These properties will be familiar to catalogers. (Source: BIBFRAME documentation)



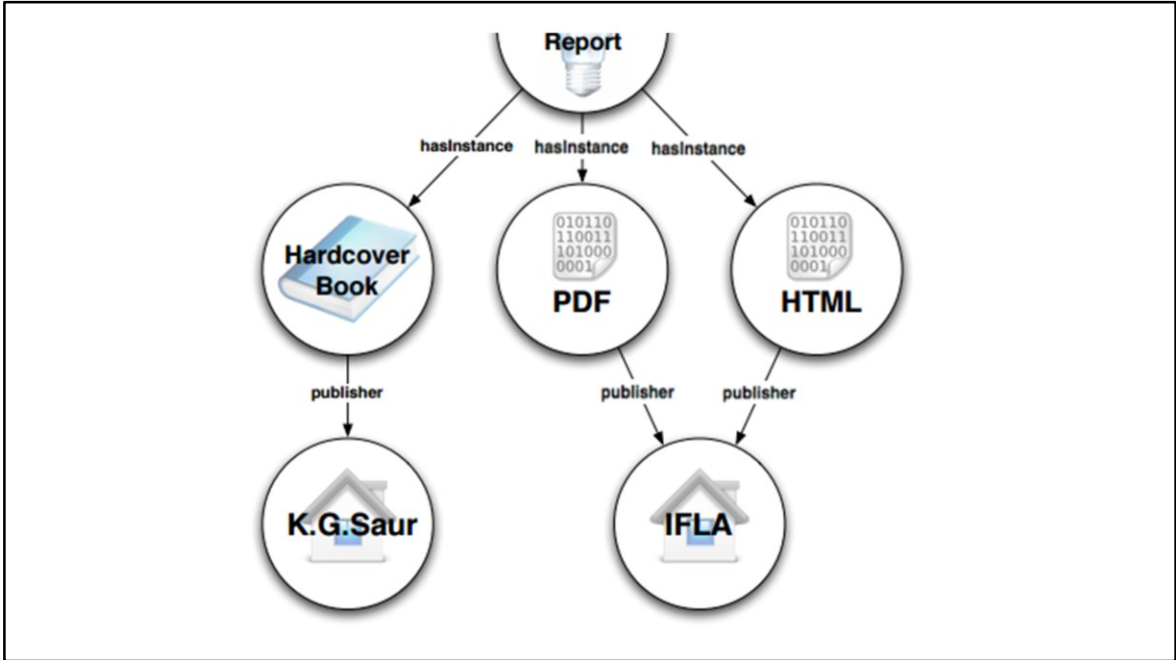
Visualizations help to see how the four core classes are related. As with FRBR and RDA, relationships are key and will allow linking on the web. In this diagram, you see a work at top center. There is a link from the Work to an Instance. There are relationships between Works and Instances, and Authorities.

Enlarged on next slides.

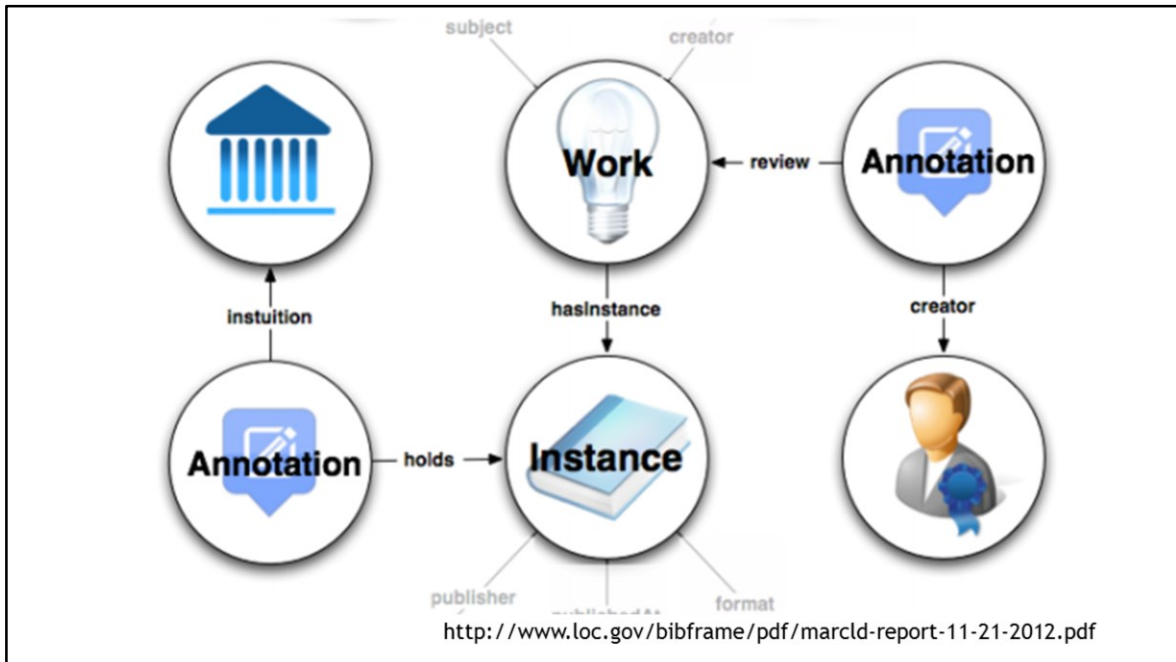
Instances as defined by this XML serialization is shown in Figure 5.



A search for “Cataloging” will lead the searcher to the Work, FRBR report.



This slide shows relationships from Instances (Hardcover book, pdf, and html versions) to Authorities (publishers). A search for IFLA will lead the searcher to Instances of the FRBR Report published by IFLA. So you can see how these Authorities can lead to bibliographic information that leads to the resources in our libraries.



This slide shows possible Annotations and related Authorities. You find information about the Work and Instance, and you will have access to other useful information: book reviews, cover art, etc.

Question: If our information is on the web, will we need OPACS? Maybe not. We'll just need web ready data and search mechanisms.

MARC fields with multiple data elements

264 1 \$a Chicago : \$b ALA Editions, an imprint of the
American Library Association, \$c 2015.

Date elements

Date of publication = 2015
Publisher = American Library Association
Place of publication = Chicago



As part of the change, we won't have fields with multiple data elements. Each data element will stand alone. And will be expressed as a triple, as Sandy explained.

What are the general differences between MARC and BIBFRAME?

(<http://www.loc.gov/bibframe/faqs/#q03>)

As a bibliographic description format, the MARC format focuses on catalog records that are independently understandable. MARC aggregates information about the conceptual work and its physical carrier and uses strings for identifiers such as personal names, corporate name, subjects, etc. that have value outside the record itself.

Instead of bundling everything neatly as a "record" and potentially duplicating information across multiple records, the BIBFRAME Model relies heavily on relationships between resources (Work-to-Work relationships; Work-to-Instance relationships; Work-to-Authority relationships). It manages this by using controlled identifiers for things (people, places, languages, etc). ... In short, the BIBFRAME Model is the library community's formal entry point for becoming part of a much larger web of data, where the links between things are paramount.



<http://www.niso.org/news/events/2013/dcmi/bibframework>

Zepheira uses this slide in their training material. It shows different objects made from Lego blocks. The idea here is that we will be able to put our data elements together in different ways, to meet various user needs. The data elements are building blocks, like Lego blocks.

BIBFRAME Vocabulary

Defines a set of classes and properties

- Class: main categorization
- Properties
 - belong to classes
 - used to describe BIBFRAME resources

I introduced the four main classes. There are about 53 classes. As BIBFRAME is being developed, more may be added.

Those classes and their properties are defined in the BIBFRAME vocabularies. In cataloging, we are used to discussing controlled vocabularies. We mean the controlled terms or values for subject headings, etc. In BIBFRAME, Vocabulary has a different meaning. The BIBFRAME Vocabulary defines the names and properties of data elements.

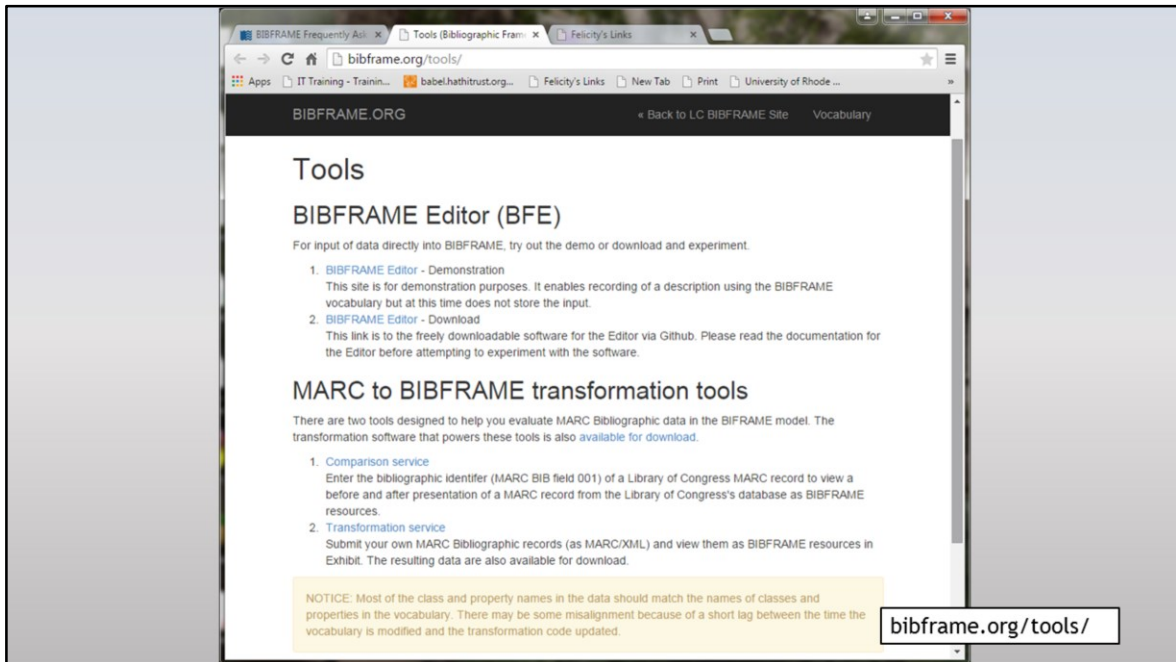
Argh! Will I have to learn rdf, triples, and BIBFRAME classes and properties to make this work?

BIBFRAME Profiles

BIBFRAME tools

- <http://bibframe.org/tools/>

Seems a bit overwhelming to have to learn computer programming lingo to do your cataloging. Don't worry. When cataloging, you'll use BIBFRAME Profiles and tools.



The Library of Congress has developed a BIBFRAME editor and MARC to BIBFRAME transformation tools. Zepheira has developed these, too. Both are being used in BIBFRAME experimentation.

BIBFRAME Editor

Get the code [here](#). In the meantime, select a Profile below to get started.

BIBFRAME - Kitchen Sink Profiles

[New HeldItem](#)

[New Instance](#)

[New Work](#)

[New Work, Instance, & HeldItem](#)

BIBFRAME - Simple Monograph

[New Holding](#)

[New Instance](#)

[New Work](#)

[New monograph \(New Work, Instance, and Holding\)](#)

Work

Authorized access point

Simple label

Work title

Any title

Title variation

Associated agent

Creator role

BIBFRAME - Simple Monograph

[New Holding](#)

[New Instance](#)

[New Work](#)

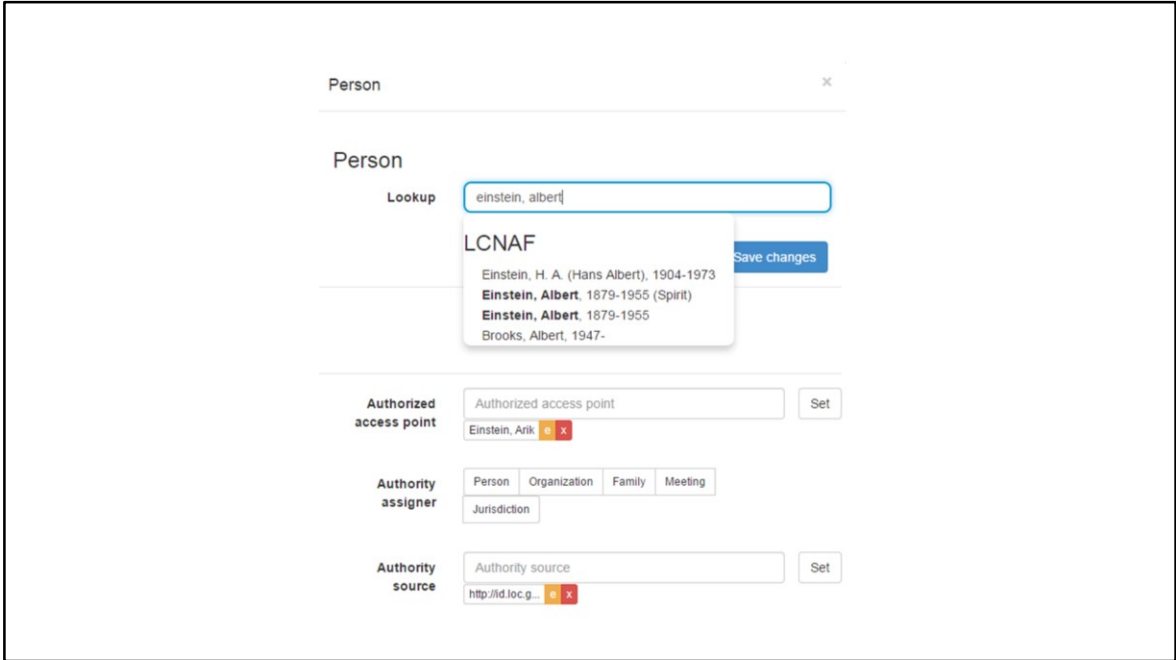
[New monograph \(New Work, Instance, and Holding\)](#)

In the BIBFRAME Editor, where you will create and, perhaps, edit records, you'll find Profiles. Profile are templates for entering data elements. (They look like records, but the data are stored as data.) Note the template for a new monograph.

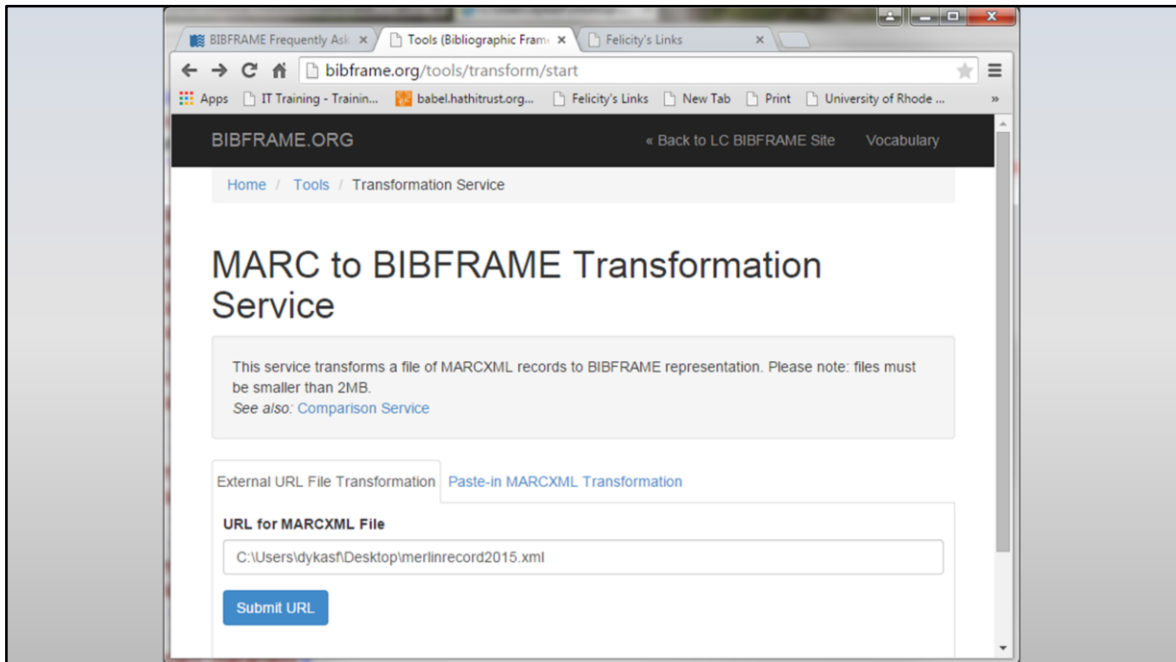
Work - Monograph

Authorized access point	<input type="text" value="Authorized access point"/>	<input type="button" value="Set"/>
Title	<input type="text" value="Title"/>	<input type="button" value="Set"/>
Title variation	<input type="text" value="Title Entity"/>	
Author	<input type="text" value="Person Organization Family Meeting Jurisdiction"/>	
Editor	<input type="text" value="Person Organization Family Meeting Jurisdiction"/>	
Translator	<input type="text" value="Person Organization Family Meeting Jurisdiction"/>	
Date of work	<input type="text" value="Date of work"/>	<input type="button" value="Set"/>
Place of creation	<input type="text" value="Place"/>	

The template includes fill-in-the blank fields and lookup fields.



Here is a pop-up box that allows one to search for a name authority record. The information can be added to the Profile from this page. So, as you can see, you don't need to know the BIBFRAME Vocabulary to do this.



Tools are also being developed to transform MARC to BIBFRAME and will be developed to transform BIBFRAME to MARC. We have a lot of MARC records that we will have to transform. Plus, as we experiment with creating “native” or original BIBFRAME records, we’ll need a way to add them to our cache of MARC records during this transition period.

Examples of Transformations

- **505 0** \$a What is RDA? -- RDA and the international context - FRBR and FRAD in RDA -- Continuity with AACR2 -- Where do we see changes? -- Implementing RDA -- Advantages, present and future.
- **bf:contentsNote** "What is RDA? -- RDA and the international context -- FRBR and FRAD in RDA -- Continuity with AACR2 - Where do we see changes? -- Implementing RDA - Advantages, present and future." ;
- **300** \$a vii, 117 p. : \$b ill. ; \$c 28 cm.
- **bf:dimensions** "28 cm." ;

This shows a simple transformation of MARC to BIBFRAME. Data for access points is more complicated, with links to authority records, etc. (Next slide)

```

bf:classification <http://id.loc.gov/authorities/classification/2094.15.04
7> ;
bf:contributor [ a bf:Organization ;
    bf:authorizedAccessPoint "American Library Association." ;
    bf:hasAuthority [ a madsrdf:Authority ;
        madsrdf:authoritativeLabel "American Library Association." ] ;
    bf:label "American Library Association." ] ;
bf:creator [ a bf:Person ;
    bf:authorizedAccessPoint "Oliver, Chris, 1951-" ;
    bf:hasAuthority [ a madsrdf:Authority ;
        madsrdf:authoritativeLabel "Oliver, Chris, 1951-" ] ;
    bf:label "Oliver, Chris, 1951-" ] ;
bf:derivedFrom <http://id.loc.gov//resources/bibs/16266874.marcxml.xml> ;
bf:hasAnnotation [ a bf:Annotation ;
    bf:annotates <http://id.loc.gov//resources/bibs/16266874> ;
    bf:changeDate "2011-02-02T15:44" ;
    bf:derivedFrom <http://id.loc.gov//resources/bibs/16266874.marcxml.xml
1> .

```

This is behind the scenes, and to catalog you will not need to know about triples, rdf, rdfa, etc.

==--==

That was a quick overview of the main pieces of BIBFRAME.

Schema.org

- Standard for semantic markup of web data
- Developed by Yahoo!, Google, Bing, and Yandex
- Structures data on the Web

- Schema Bib Extend Community Group
 - <https://www.w3.org/community/schemabibex/>

- Overlap with and complements BIBFRAME

Quick introduction to Schema.org. BIBFRAME is not the only possible schema for coding web data. Schema.org was developed by Yahoo!, Google, Bing, Yandex. It is being used in industry, commercial ventures, etc. It is one reason you see more and more structured data in search results when you search on the web. OCLC has coded bibliographic data from WorldCat using Schema.org. Discussions by the Library of Congress, OCLC, and others have led to the conclusion that Schema.org and BIBFRAME are complementary. Schema.org does not cover the needs of bibliographic descriptions as well as BIBFRAME.

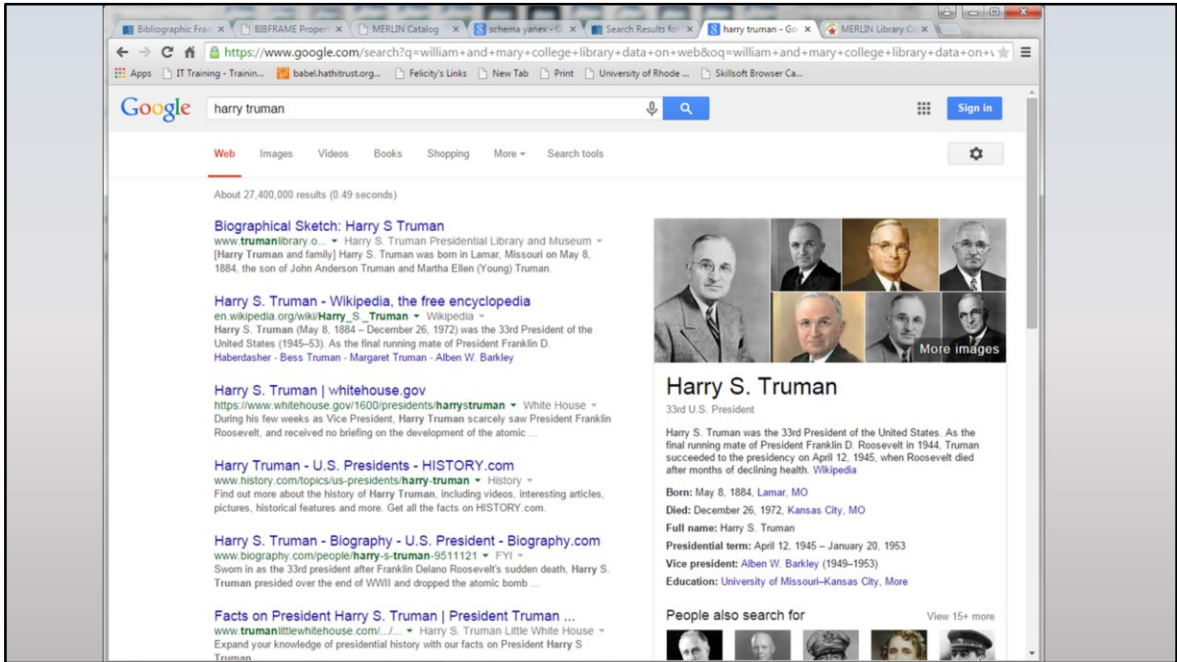
February 2015 update

- Year of experimentation completed
- Next steps
 - Adjust BIBFRAME vocabulary
 - Continued development of tools
 - Pilots for original cataloging
 - Library of Congress
 - Stanford and five other libraries
 - Further development of MARC to BIBFRAME and BIBFRAME to MARC conversion tools
 - OCLC: more work with OCLC Schema.org and BIBFRAME approaches
 - Zepheira working on tools and working libraries on projects
 - Large scale experimentation

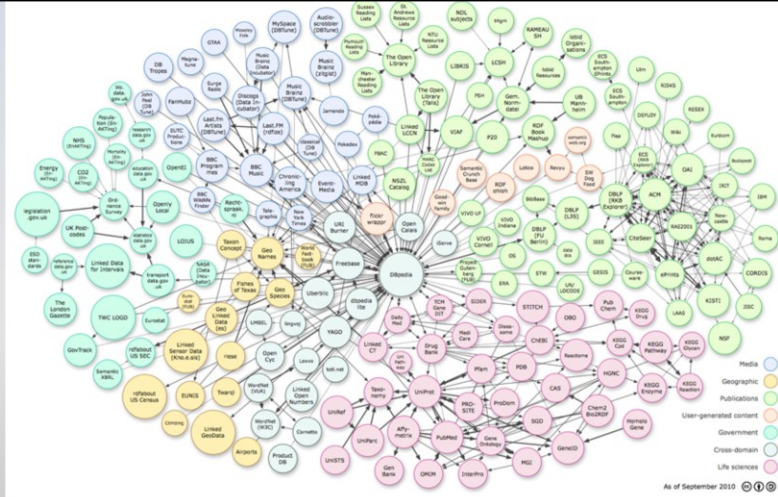
http://www.loc.gov/today/cyberlc/feature_wdesc.php?rec=6647

We can expect changes to BIBFRAME.

- Adjust vocabularies: Develop principles, Consistency, Expansions
- Library of Congress pilot: Original cataloging in BIBFRAME. Still developing needed tools (e.g., BIBFRAME editor). Will share tools.
- Stanford and five other libraries have pilot to catalog using BIBFRAME: Will learn more about authorities and resource files, different workflows, etc.
- Recent posts to the BIBFRAME email list included discussions about: triplestores, triples, datastores, OWL, SPARQL queries, if the endpoint is a literal, serialization, RDF vocabularies, graph properties, computational ontologies, inverse properties. So many details are still be discussed and BIBFRAME is being refined.



As I conclude this section, a reminder about the goal: get our bibliographic data on the web. Currently, bibliographic information from our OPACS/ILS are not available in search results. People are not finding the relevant resources we have.



Show Me the Linked Data in Libraries!

My interest in linked data has led me to attend quite a few presentations and webinars in the past years, filling my head with images such as this Linking Open Data cloud and visual metaphors for linked data. The promise that linked data gives us is that all the careful metadata work we've done over the years will usher us into a new era of visibility on the web where our users our. I'm convinced & ready to move on to this shiny new world! Now what is the progress being made? Show me the Linked Data in libraries. My part of the presentation will focus on what linked data projects and developments are happening in US libraries, as otherwise it could easily take hours to cover all the European projects.

BIBFLOW

AN IMLS PROJECT OF THE UC DAVIS LIBRARY & ZEPHEIRA

[HOME](#) [ABOUT](#) [RESEARCH QUESTIONS](#) [FINDINGS](#) [REPORTS](#) [TEAM](#) [CONTACT](#) [WORKFLOW ANALYSIS](#)

- A two year IMLS grant project of the University of California Davis Library & Zepheira
- Goal: Switch entire database to graph-based/triple-store
- Underlying model: BIBFRAME
- Software: Kuali Ole (with coding modifications)
- Seeking feedback and comments to make model as universal as possible
 - Use cases

First, BIBFLOW provides a good transition from Felicity's presentation on BIBFRAME. This is a 2 year grant-funded project awarded to the University of California-Davis and Zepheira. I've placed a link to a video on their website that does a great job of explaining the project. As the presenter explains, it's not reasonable to convert to linked data right now because catalogs are more than just containers for bibliographic data, they have specialized functions, such as circulation and interlibrary loan. This project is an attempt to move us toward that eventual migration.

They will be working toward switching the entire database to a native triple-store. In other words, not just crosswalk from MARC but actually create and store data in triples. They will use BIBFRAME as the underlying model and they will modify an open source software called Kuali Ole. Ole itself was a grant project and it stands for the "Open Linked Environment."

Copy Cataloging for Non-Rare Books Using OCLC Connexion Client

posted on October 25, 2014 by Carl G Stahmer | 2 Comments

The following workflow represents the current work process in place at the UC Davis library when copy cataloging a non-rare book using OCLC connection client. It is one of the work functions that will be converted to a native Linked Data workflow as part of the project.

Actors

1. Cataloging Specialist – Medium level of experience.
2. Copy Cataloger – Lowest level of experience.
3. Supporting systems and tools (e.g. harvesting daemon).

Description

User Story: "As a library cataloger I need to find and claim a record in OCLC for a published book to include in my integrated library system with additional metadata for its location in library collection."

Success Metric(s)

1. The new set of bibliographic metadata elements will be present and discoverable by library users of the OCLC Melvyl System.

If you go to their website, you can see that they are documenting use cases. For example, here is an use case about cataloging monographs using OCLC Connexion. Your use case might differ, particularly if you share records with other institutions, such as with MOBIUS. They have emphasized that they want this model to be as universal as possible, not just what works for UC-Davis, which is why they're inviting other libraries to contribute how their workflows differ.



"I believe everyone benefits from the visibility of libraries and their content on the Web."

- The Libhub Initiative Pledge

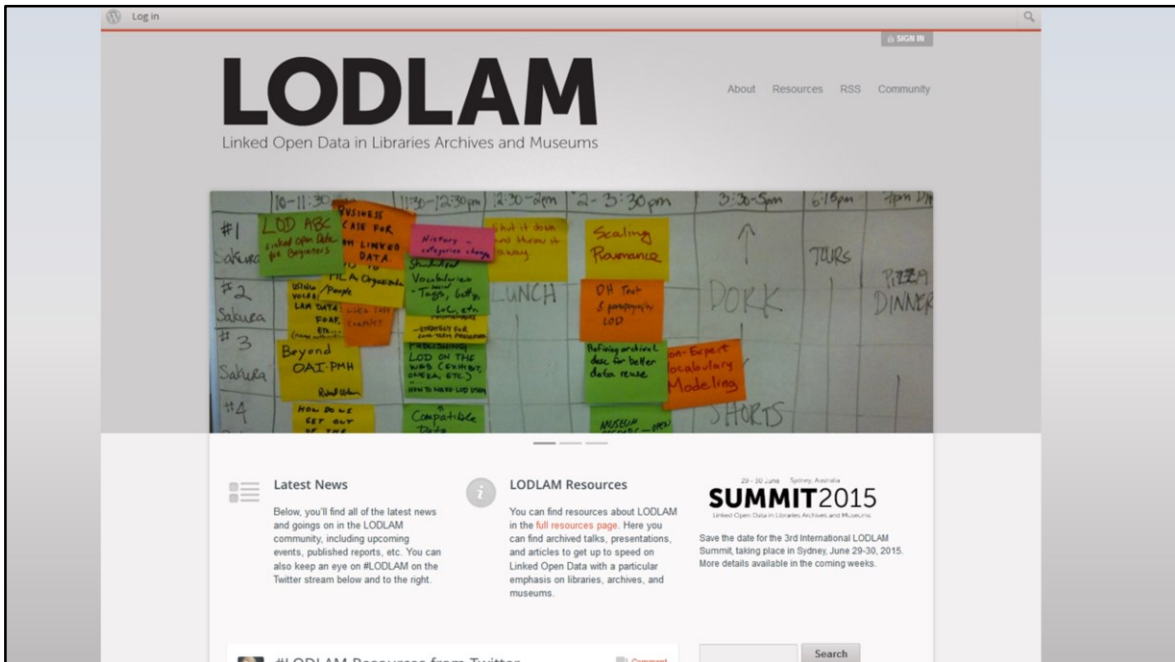
- One of founding sponsors: Zepheira
- Other vendor sponsors: Atlas, Innovative, SirsiDynix
- Goal: "publish BIBFRAME resources to the Web" and enhance discoverability

LibHub is another project, and notably Zepheira is listed as the founder. This project is aiming to build momentum for the goal of "publishing BIBFRAME resources to the web". In contrast to BIBFLOW, they expect that libraries will catalog using MARC for years to come, so in the meantime they're proposing to crosswalk the data to BIBFRAME. What's notable is that ILS corporations have signed up, most notably Innovative. The truth is that most libraries are dependent on vendors for change, so the big question is if and when they'll be able to implement BIBFRAME in some form.

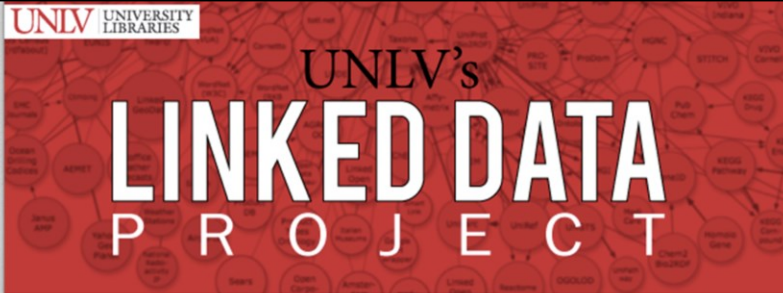
Linked Data for Libraries (LD4L)

- Collaboration of Cornell, Harvard and Stanford
- Funded by nearly \$1 million two-year grant (Andrew W. Mellon Foundation) in February 2014
- Goal of project: Create a triple store called the Scholarly Resource Semantic Information Store (SRSIS) model
 - Intent to use existing ontologies
 - Also to use open source software
 - One deliverable: to build a Blacklight interface to facilitate searching of triples

Next, Linked Data for Libraries, LD4L, is another 2 year grant project between 3 big universities: Cornell, Harvard, & Stanford. The scope is metadata within the realm of libraries, such as MARC records in the catalog, repository, and digital library metadata. They are working to create a triple-store called the Scholarly Resource Semantic Information Store. They will follow recommended practices and not re-invent the wheel, but instead re-use existing ontologies, and that includes BIBFRAME for MARC records. They will also use open source software. A stated deliverable is to build a Blacklight interface that has a more user-friendly interface to search triples. Right now, you can use SPARQL to search triples, but not without being pretty technologically savvy yourself and you wouldn't expect your users to query the database with it.



Finally, there's LODLAM. LODLAM has more grassroots origins, per its webpage that "LODLAM.net is an informal, borderless network of enthusiasts, technicians, professionals and any number of other people who are interested in or working with Linked Open Data pertaining to galleries, libraries, archives, and museums." They host conferences called LODLAM Summit. This year's is taking place in Sydney, but before you pack your bags, this one is invitational only and limited to 100 people who work actively with linked data or have the authority or influence to bring about linked data.



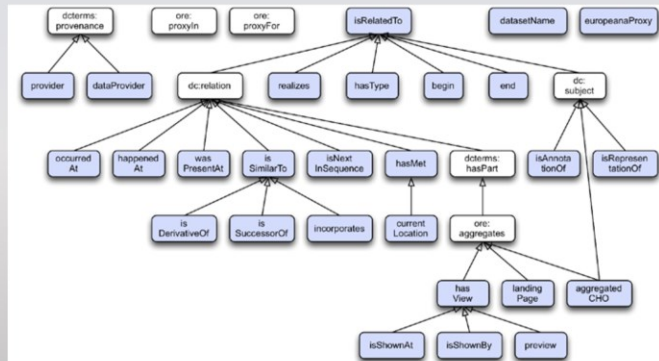
- Inspiring because presenters have emphasized that their “Project led, implemented & managed by 2 busy faculty librarians”
- Library first formed a study group on linked data
- Then after a year, worked on prototype
 - Enabled them to verify the ability to generate LOD without investment in new systems
 - Instrumental in acquiring administrative support for project

Finally, there’s the University of Nevada-Las Vegas linked data project. I found myself enthusiastic about this project because of this quote from one of their presentation slides: “Project led, implemented, & managed by 2 busy faculty librarians.” Bingo, I’m busy and a faculty librarian: is this something I could work on? Plus, they are closer to the funding and my type of library than say, Harvard or Stanford. What follows is a summary from a few presentations I’ve watched as well some articles they’ve published, all included in the links.

They didn’t jump into a linked data project right away. Instead, they started a study group of interested people in their library to explore this concept and see what all the excitement was about. After a year, they felt confident enough to start a prototype project, that addressed two goals. One: to prove that they could actually implement linked data without investing in new software and two: to gain the needed support from their administration.

Design Phase

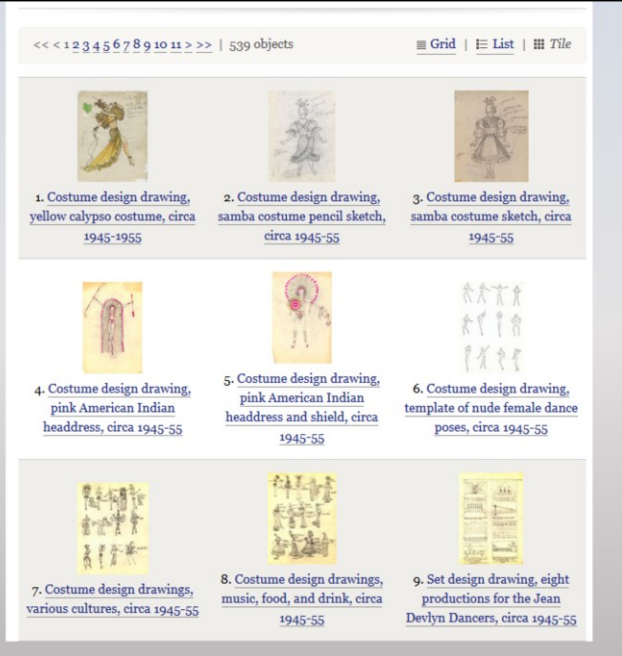
- Selecting Technologies
- Data model
- Mapping
- Determining rules to create URI's




Then followed the design phase, where they put all they had learned the previous year to work. This included choosing the software they would use, evaluating and modifying the data model they would use. The data model they used was the Europeana Data Model, with a few modifications. Then there was the mapping, and choosing where they would obtain their URI's, which would form the basis for linked data.

Implementation

Initial Data prep in
ContentDM repository



Now finally we get to the implementation phase. Their collection was housed in ContentDM and a screenshot of it is on the right-hand of the slide here. The first step, as they learned through experience, was to do as much data prep in ContentDM as they could. Then, once they were satisfied, they exported this data and imported it into a program called Open Refine.



Refine *OPEN* *A free, open source, powerful tool for working with messy data*

- Further Metadata Work
- Reconciled Data
- Generated triples/RDF schema & exported theses

Google refine masters_theses.xls_Thu_Apr_23_06_19_44_2015.xls Permalink

Facet / Filter Undo / Redo

6713 rows

Show as: rows records Show: 5 10 25 50 rows

All	title	title_alternative	fulltext_url	keywords	abstract	advisors_list	committee_me	disciplines	department
1.	An installation and migration methodology for operating systems							Computer Sciences	Computer Science
2.	An experimental numerical							Engineering Mechanics	Mechanical and Aerospace Engineering

Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

This is a program that looks very much like a spreadsheet program, as you can see from a screenshot at the bottom of the screen. But in reality it is so much more. It is able to cluster data using facets and filters and help you quickly and easily clean up the metadata. It is a great tool for cleaning up “messy” data as the slogan says. It can also retrieve an URI from linked data stores, such as if you have a library of congress subject heading, it can retrieve that URI from the Library of Congress site. It also has an extension called RDF extension that can form the triples.

Import linked data into triple stores

- Runs parallel to their ContentDM repository
- Finished result: Video of content in Pivot Viewer:
 - <https://youtu.be/-83FTKEkYZ0>

From Open Refine, they exported the triple into a triple store. They have a very attractive video on their site showing the result in Pivot Viewer. They eventually intend to have both systems display in tandem.

How Feasible is it to Implement Linked Data?

- Independent project requires
 - Technical knowledge or willingness to learn (RDF, linked data principles)
 - Group project
 - Administrative support/buy-in
 - IT/Technical Support (triple store)

So bottom line: what can we do to implement a linked data project? First, as you may have gathered, it will take a willingness to learn technical subjects such as RDF and triples: they didn't teach this years and years ago in my library school! Also, the more you can collaborate with colleagues, the better. Getting buy-in from your administration by demonstrating that the benefits and potential return of investment is also necessary. Finally, you will need IT to install software, such as triple stores.

When Can We Expect to See Linked Data in Our Libraries?

- Pioneers/early adopters are paving the way
- Not financially feasible to leap until details have been worked out
- Momentum may propel linked data era forwards

So, when can we expect to see linked data in a library near you, or even our own libraries? As the presenter in the BIBFLOW video says, we aren't yet ready for prime time. And this is what the grant-funded projects and the early adopters in libraries are doing: moving us forward, helping us to learn what is practical and what isn't. The more information we have, the more we can make informed decisions and eventually our ILS systems will implement this. How fast this will happen is up for debate—it seems likely it will take many years for the library world to transition, on the other hand we don't know what technologies or software is on the horizon and all this momentum may bring it about faster than we anticipate. Stay tuned for further developments.

References

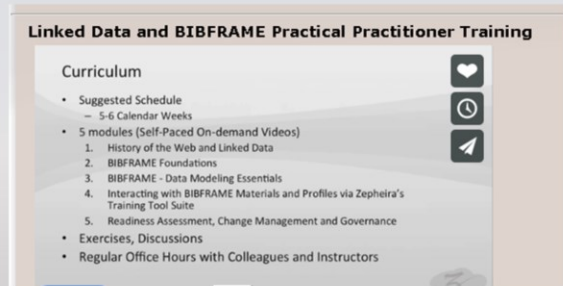
Will be posted on the MOBIUS Conference Website

Questions

- Humorous Disclaimers:
 - We're 3 busy faculty librarians
 - We're librarians, not "linked data specialists," Jim
 - Our crystal balls work imperfectly



Zepheira Training



Linked Data and BIBFRAME Practical Practitioner Training

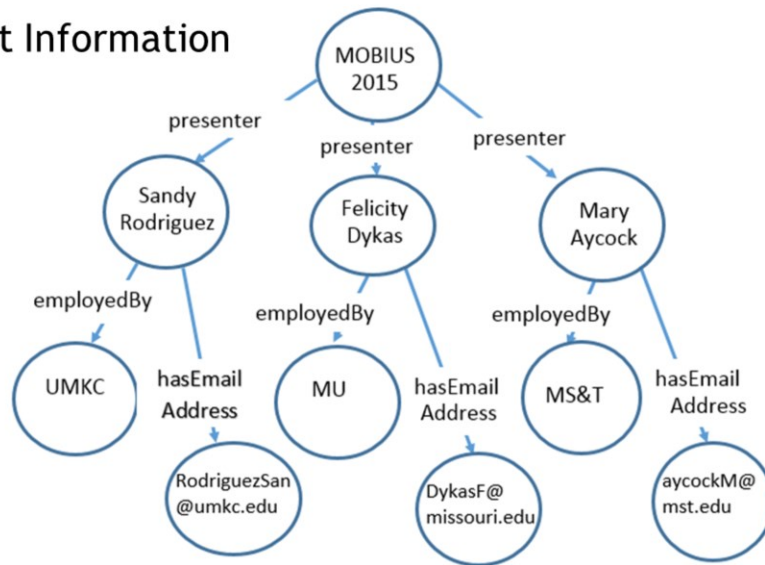
Curriculum

- Suggested Schedule
 - 5-6 Calendar Weeks
- 5 modules (Self-Paced On-demand Videos)
 1. History of the Web and Linked Data
 2. BIBFRAME Foundations
 3. BIBFRAME - Data Modeling Essentials
 4. Interacting with BIBFRAME Materials and Profiles via Zepheira's Training Tool Suite
 5. Readiness Assessment, Change Management and Governance
- Exercises, Discussions
- Regular Office Hours with Colleagues and Instructors

On the right side of the curriculum list, there are three icons: a heart, a clock, and a paper plane.

<https://zepheira.com/solutions/library/>

Contact Information



Sandy Rodriguez, University of Missouri—Kansas City, RodriguezSan@umkc.edu

Felicity Dykas, University of Missouri—Columbia, DykasF@Missouri.edu

Mary Aycock, Missouri University of Science and Technology, AycockM@mst.edu