BIBFRAME for Dummies: What Can Catalogers Do Now?

NOTSL Fall Meeting, October 30, 2015
Cuyahoga County Public Library
Parma, OH
by
Roman S. Panchyshyn
Catalog Librarian, Assistant Professor
Kent State University Libraries
Introduction

This presentation will address these questions

• What can catalogers do now to prepare for BIBFRAME (BF)

• Some insight into what role will cataloging staff play in a future BF environment
Areas of Concentration

• Learn the vocabulary
• Who are the players/projects
• What are the tools out there
• Can I participate?
• Future scenarios
What is BIBFRAME (BF)

• Web of links
• Linked data model to replace MARC
• Will deconstruct MARC data and replace it with linkable information resources
Four Main Classes

• Creative Work – conceptual essence of the cataloging item.
• Instance – reflects an individual, material embodiment of the Work.
• Authority – resource reflecting key authority concepts that define relationships reflected in the Work and Instance.
• Annotation – resource that decorates other BF resources with additional information; such as library holdings, cover art, reviews.
Library of Congress BIBFRAME MODEL
RDF

- Resource Description Framework
- Semantic web standard
- Standard model for data interchange on the web
- Describes a graph database
Bengie is a dog.
Bonnie is a cat.
Bengie and Bonnie are friends.

Simple data graph with properties
Triples

• RDF statements are called triples
• Have subject, object, predicate

Subject is the T-shirt
Predicate (property) is the color
Object is white
RFF in XML

```
01. <xml version="1.0" encoding="UTF-8"/>
02.
03. <rdf:RDF
04.     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
05.     xmlns:feature="http://www.linkeddatatools.com/clothing-features#"/>
06.
07.   <rdf:Description rdf:about="http://www.linkeddatatools.com/clothes#t-shirt">
08.     <feature:color rdf:resource="http://www.linkeddatatools.com/colors#white"/>
09.   </rdf:Description>
10.
11. </rdf:RDF>
```

T-shirt statement in RDF
Key Point to Remember

• A subject in an RDF document may also be referenced as a object of a property in another RDF statement
• Unique IDs can be Uniform Resource Identifiers (URI)
Graph with URIs

These URIs can be links to authority records
Registered Vocabularies and Extensions

• In the linked data environment, for computers to communicate with each other, and with search engines, they need to share a common vocabulary.

• More formal vocabularies can also be referred to as ontologies.
Schema.org

- Launched in 2011 by Bing, Google, Yahoo
- Allows people to create and support a common set of schemas for structured data markup on web pages.
- RDF is one such schema
Extensions

- Schema.org allows users to create extensions to its vocabularies
- Registered vocabularies exist for Dublin Core, RDA, and BF
- These can all be extended to RDF
BF Vocabulary

• Zepheira and LC worked together to create the BF vocabulary, found here: http://bibframe.org/vocab/

• BF Vocabulary is comprised of the RDF properties, classes, and relationships between and among them

• Computers can now share a common BF vocabulary
This example: RDF points to BF vocabulary

```xml
<?xml version=\"1.0\" encoding=\"UTF-8\"?>
<rdf:RDF
  xmlns:bf=\"http://bibframe.org/vocab/\"
  xmlns:bfp=\"http://bibframe.proposed.org/\"
  xmlns:madsrdf=\"http://www.loc.gov/standards/mads/rdf/\"
  xmlns:rdf=\"http://www.w3.org/1999/02/22-rdf-syntax-ns\">
  <rdf:Description rdf:about=\"http://viaf.org/viaf/59083797\">  
    <bf:label>Herbert, Frank</bf:label>  
    <madsrdf:isIdentifiedbyAuthority rdf:resource=\"http://id.loc.gov/authorities/names/n80044450\"/>  
    <rdf:type rdf:resource=\"http://bibframe.org/vocab/Person\"/>  
  </rdf:Description>
</rdf:RDF>

<rdf:Description rdf:about=\"http://herbert.example/4\">  
  <bf:isbn>0801950775</bf:isbn>  
  <bf:isInstanceOf rdf:resource=\"http://herbert.example/2\"/>  
  <bf:date>1965</bf:date>  
  <bfp:barcode>1965</bfp:barcode>  
  <rdf:type rdf:resource=\"http://bibframe.proposed.org/Hardcover\"/>
</rdf:Description>
```
Searching

- For linked data to be useful for humans, we have to develop ways for searching it.
- Searching needs to be done within RDF
SPARQL

• SPARQL (a recursive acronym for SPARQL Protocol and RDF Query Language) was developed as an RDF query language

• This is a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF)
Databases

- Linked data must be stored as triples
- Databases/servers that do this are referred to as “triple stores” or “3store”.
- Most ILS databases do not yet have this capability.
Semantic Web

• The vocabulary that has been described here is broadly referred to as the semantic web.

• This a common framework that allows data to be shared and reused across application, enterprise, and community boundaries

• Key: data must be open
The Players

• This section will examine some of the key players and projects that are currently underway with BF
• List is non-exhaustive, North American centered
Library of Congress

• Worked with Zepheira to develop the BF vocabulary
• Maintain documentation for the project
  – http://www.loc.gov/bibframe/ (main page)
  – http://bibframe.org/ (technical site)
• Developing tools and training to be shared by libraries
• Currently pilot testing a BF workflow
Zepheira

- Private corporation, founded by Eric Miller
- Based in Ohio
- Contracted by LC to develop BF Initiative
- Leading provider of linked data and BF training
- Located at: [http://zepheira.com/](http://zepheira.com/)
Zepheira’s Libhub Initiative

• Aims to convert hundreds of library’s bibliographic records and publish them on the Web in BF to build a core set of library data on the Web

• Persons using a search engine would find library data and be then taken to the local OPAC or discovery layer.
Libhub 2

- Costs associated with project (subscription)
- Zepheira will convert library database to BF
- Zepheira will maintain 3store database
- Zepheira will provide library training
Some Libhub Participants
OCLC

- Host most of the bibliographic data we use
- Active in linked data research, have been developing their own linked data model, the OCLC/Schema model in contrast with the BF model
- Supports VIAF (The Virtual International Authority File), an international service providing access to the world's major name authority files
Linked Data for Libraries

- Linked Data for Libraries Project (LD4L)
- Funded by $1 million two-year grant from the Andrew W. Mellon Foundation.
• Goal is to create a Scholarly Resource Semantic Information Store (SRSIS) model that works both within individual institutions and through a coordinated, extensible network of Linked Open Data
• Capture the intellectual value added by librarians and other domain experts and scholars when they describe, annotate, organize, select, and use those resources
• Found at: https://www.ld4l.org/
National Library of Medicine (NLM)

- Early BF experimenter
- Broke off with LC in 2014 to work with Zepheira, George Washington University (GWU), and University of California, Davis (UCD) in development of the BF Lite vocabulary, as hosted by Zepheira
NLM

• Mapped the PCC RDA BIBCO Standard Record Metadata Application Profile (BSR, as of April 14, 2015), BF Lite (as of June 8, 2015), and RDA RDF (as of June 23, 2015) properties.

• Focusing workflow on creating new cataloging data directly in BF rather than converting legacy bibliographic data.
BIBFLOW

• IMLS project between UC Davis and Zepheira
• Goal is to investigate the future of library technical services (cataloging and related workflows) in light of modern technology infrastructure and new data models and formats such as Resource Description and Access (RDA) and BIBFRAME
Libraries currently constrained by complex workflows and interdependencies on a large ecosystem of data, software and service providers that are change resistant and motivated to continue with the current library standards.

Information found at: https://www.lib.ucdavis.edu/bibflow/
Other BF Implementers and Testers

- Any library that is testing or implementing BF is asked to add their name to the BIBFRAME Implementation Register page
- Page located at: http://www.loc.gov/bibframe/implementation/register.html
BF Tools

- This section will recap what tools are out there for library staff to use to familiarize themselves with BF and BF implementation.
- 3 main sources of tools for experimenting:
  - Library of Congress
  - Zepheira
  - MarcEdit
LC Tools

• Located:  
  http://www.loc.gov/bibframe/tools/

• Most important
  – Metaproxy X-Query
  – Metaproxy SPARQL
  – BIBFRAME Editor
  – Comparison Service (MARCXML to BF)
  – Transformation Service (MARCXML to BF)
LC Training

• LC also developing training modules for staff as part of their pilot, that are being shared freely. These are still in process.

• See: http://www.loc.gov/catworkshop/bibframe/
Zepheira BF Tool

- BIBFRAME Scribe prototype
- Found at: http://editor.bibframe.zepheira.com/static/index.html
- Demonstrates how to catalog various materials natively in Linked Data, being modified with support from UC Davis University Library as part of the BIBFLOW project
BF Scribe

- Modular in nature
- Choose instance (book, e-serial, etc.)
- Fill in the information in appropriate sections
- Option to save or export completed record in RDA/BF Lite
BF Scribe

• External links to:
  – Library of Congress Linked Data Service (names, subjects, languages, places, RDA categories)
  – assignFAST (subjects)
  – VIAF (names, subjects)
  – AGROVOC (subjects)
  – Medical Subject Headings (MeSH) RDF Linked Data (subjects)
Once you are done, you can save or export in RDF/XML
BF Scribe

- Tool is modular, catalogers will not see back end operations, only enter data
- Tool still under development
MarcEdit

- Developed and maintained by Terry Reese, The Ohio State University
- Found at: [http://marcedit.reeset.net/](http://marcedit.reeset.net/)
- Section on MarcNext contains BF tools for testing
- Working with Zepheira to get these tools integrated into BF workflow
The BibFrame Testbed is a simple to use utility designed to allow metadata specialists the ability to model their data using BibFrame Concepts.

Source File: 

Save File: 

File Type: 

BaseURI: http://example.com/ 

XQuery URI: http://marcedit.reerset.net/software/xquery Edit 

Data Serialization: RDFXML 

Allows you to model data using BF concepts
JSON view of OhioLINK record: [http://olc1.ohiolink.edu/record=b19807580~S0](http://olc1.ohiolink.edu/record=b19807580~S0)
The Linked Data Tool has been designed to loop through a set of MARC records and resolve access points in the 1xx, 6xx, and 7xx data elements to their linked data end-points.

Resolve access points
Before
Links to LC and VIAF added in $0 in 1XX and 6XX tags
SPARQL Browser

SPARQL Browser

SPARQL Endpoint:

SPARQL Query:

Output: XML

Results:

Query RDF Databases

KENT STATE
University Libraries
MarcEdit Summary

• Powerful, useful tools
• Allows TS staff to take more control of the process, especially the linked data tool
• Adapting local workflows to add $0 to headings insures that the URI is in place when converted to BF
OCLC FAST Linked Data

- FAST (Faceted Application of Subject Terminology)
- Based on deconstruction of Library of Congress Headings
- All authority records available free to download
OCLC Fast Tools

**FAST Projects**

- **searchFAST**
  A full feature search interface to the FAST database.

- **mapFAST**
  mapFAST is a Google Maps mashup prototype designed to provide map based access to bibliographic records using FAST geographic and event authorities. Also available in Mobile versions.

- **FAST Converter**
  The FAST Converter is a Web interface for the conversion of LCSH headings to FAST headings.

- **FAST Linked Data**
  FAST as a Linked Data service to interact with the Semantic Web.

- **assignFAST**
  A Web service that automates the manual selection of FAST headings.
LCSH and FAST in OCLC Record
BF Participation

• If your library is interested in BF participation on your own, you will need:
  – Tools
  – Training
  – Staff resources
  – Hardware/software support

• You also have the option to partner with Zepheira
Tools

• Free tools available, but not yet at the scale that library may need to convert all data and workflows to BF.
• Pilot tests not yet complete
Training

• LC training on linked data/semantic web not yet complete
• Zepheira can provide staff training, but at cost
• If tools are modular, how much does staff really need to know about back end?
Staff Resources

• How much does your staff know about linked data?
• New skill sets need to be developed in department
• Programming support may be necessary for database conversions
Hardware/Software Support

• You may need to go outside of department for software support
• Data needs to be stored in a 3store database. Can your ILS support this?
Participation

- Developers from one institution can freely contact others through the BIBFRAME Implementation Register
- Can your institution provide the support you need for BF testing?
- You always have the option to work with Zepheira
BF Future: Going Forward

• There are some major challenges that need to be addressed as BF pushes forward
• The next few slides will address these challenges, in no particular order of importance
Things not Strings

- Identifiers (authorities) must be established for persons, places, objects, concepts, on a continuing basis.
- Contributions to NACO/SACO and ultimately VIAF will increase in importance.
Data Cleanup

• Some of our legacy data is dirty and will be difficult to clean up

• Essential to clean up as much of it as possible before BF conversion. Consider RDA enrichment.
Data Conversion

• Who will do legacy database conversions? Will these tools become freely available?

• How much data would we be willing to lose in a BF conversion? Vocabularies, ontologies used may not cover every bit of information in a MARC record.

• Would we archive our old MARC data?
Linked Open Data

• Linked data systems work best when the data is open and freely accessible.
• Not all bibliographic data is open. Lack of open data will impact development and services, difficult to share with other communities.
ILS Future

• ILS will be slow to move away from MARC based systems
• Require substantial commitment in resources to develop new systems/databases
Technical Services Futures

• Modular tools being developed.
• Workflows will need to be redesigned, documentation updated
• Skill sets of staff must be adjusted/updated
• How will transition to BF be evaluated?
BF Flavours

• We have various “flavours” of BF being developed
  – LC BF project
  – Zepheira BF Lite
  – OCLC Linked Data Model
• Can they coexist? Will one model emerge as the best practice?
Responsibility

• Who will take responsibility to curate the vocabularies and datasets on a national and international scale?
Conclusion

• In my opinion, we are still several years from abandoning MARC-based systems altogether
• But we need to make our data visible
• We must make the effort to make this technology work for us, and we must control the process
Questions
Contact Information

Roman S. Panchyshyn,
Catalog Librarian, Assistant Professor
Kent State University
330-672-1699
rpanchys@kent.edu