Library of Congress
BIBFRAME Progress

March 27, 2019
Presenters

Judith Cannan
Chief, Cooperative and Instructional Programs Division

Sally McCallum
Chief, Network Development and MARC Standards Office

Kevin Ford
Network Development and MARC Standards Office

Paul Frank
Cooperative and Instructional Programs Division
BIBFRAME Goals for FY19
Judith Cannan
ABA Director’s goals

• Expand the BIBFRAME 2.0 Pilot to 100 active BIBFRAME participants

• Expand the ability to import records from external sources

• Explore and evaluate Discovery as it relates to BIBFRAME

• BIBFRAME database available for public consumption
FY19 tasks

- NDMSO and COIN
  - Improve communication with LC pilot participants
  - Work with LD4P on training and collaboration on technical issues
  - Many of the following tasks are carried out jointly by NDMSO and COIN
NDMSO tasks

- BIBFRAME database
  - Convert the full MARC file again, with revised merge algorithms
  - Compare the Library’s and Casalini’s work structures for descriptions converted from MARC records
  - Enable database access by outside institutions
  - Enable import of external BIBFRAME descriptions
NDMSO tasks

• BIBFRAME Editor
  
  o Implement enhancements based on collaboration with Pilot participants, such as individualized templates, rollover information, provider lists, improved labeling, etc.
  
  o Explore more data validation
  
  o Improve BIBFRAME Profile Editor functionality
NDMSO tasks

• BIBFRAME to MARC conversion
  o Complete and test transformation
  o Identify MARC changes needed
  o Plan implementation
COIN tasks

• BIBFRAME Editor Manual

  o Create a comprehensive BIBFRAME Editor manual, which will include:

    ▪ New features as they are developed
    ▪ Guidelines and “At-a-glance” materials
    ▪ Training materials as they are created
COIN tasks

• BIBFRAME training
  o Develop and refine training materials
  o Instruct new BIBFRAME Pilot participants
COIN tasks

• BIBFRAME Pilot participants
  o Expand participant pool
  o Meet regularly with participants, in one-on-one settings, and in divisional meetings
  o Track and monitor BIBFRAME Editor requests made by participants
MARC TO / FROM BIBFRAME
Sally McCallum
MARC converted to BIBFRAME, briefly

- Eliminated much duplication
- Dropped indicators and codes that were no longer relevant
- Converted codes and data strings like names to URIs
- Differentiated “statements” (transcription) and access forms of information
- Split data into parts with emphasis on relationships
BIBFRAME to MARC conversion

- Data model issues
- Treatment of non-Latin script
- Adding URIs to MARC
- Missing punctuation
- Summary
BIBFRAME to MARC conversion

Motivations

- Double keying of bibliographic data in Pilot
- Need for “easy” transformations
  - Transformations will occur multiple times, back and forth
- Need to supply both full MARC records and full BIBFRAME descriptions into foreseeable future
- Many systems will probably need MARC for some subsystems even if they use BIBFRAME for the catalog
MARC and BIBFRAME Data Models

MARC – Unit record model
- Records include descriptions, subjects, some links, some work titles (i.e., uniform titles)
- Embeds different media in some cases – print, electronic, microform
- Embeds different physical copies in some cases – paper print, 16 mm master, 16 mm reference print, library binding, paperback, etc.

BIBFRAME – Splits information into Works (RDA work/expression), Instance (RDA manifestation), Items
- Work – conceptual focus (subjects, authorized access point, etc.)
- Instance for each media
- Items for different physical copies of an Instance
How to repackage going to MARC?

BIBFRAME Work → MARC title authority?

or

BIBFRAME Work → MARC bibliographic (work)?

What does a “MARC bibliographic work” look like?

- Contains subjects? Genre/Forms?
- Contains links to the MARC bibliographic manifestations?
- Contains Authorized Access Point
How to repackage going to MARC?

BIBFRAME Work + BIBFRAME print and electronic Instances →

MARC Bibliographic record with embedded print and electronic?

or

MARC bibliographic for print?
MARC bibliographic for electronic?
Non-Latin script data

MARC Bibliographic records use “Model A”:
- All non-Latin in 880 fields
- Fields 880 linked to corresponding regular fields that contain transliterated data
  880 10 $6 245-02/$1 $a 樊巌 集
- Transliterated data: Main entry, titles, imprints, notes, etc.

MARC Authority uses “Model B”:
- Non-Latin and transliterated information in regular fields
Non-Latin in BIBFRAME

BIBFRAME descriptions

- Transcribed data (Instance/manifestation data) is primarily only in the script on the resource
  - Title statement and statement of responsibility, edition statement, publication statement, etc. all in original script
  - Instance/manifestation description is linked to the Work and to Names where the transliterated data is found for the title, contributors, etc.

Impact on MARC bibliographic records

- Use of 880 fields for transliterated transcribed data minimized or eliminated
URLs from BIBFRAME data

- Many elements that do not normally have URLs in the MARC record will bring them from the BIBFRAME environment

**MARC data before conversion to BIBFRAME**

```
<datafield tag="700" ind1="1" ind2=" ">
  <subfield code="a">Bauer, Maria Poglitsch,</subfield>
  <subfield code="e">translator.</subfield>
</datafield>
```

**MARC data transformed from BIBFRAME**

```
<datafield tag="700" ind1="1" ind2=" ">
  <subfield code="a">Bauer, Maria Poglitsch</subfield>
  <subfield code="e">Translator</subfield>
  <subfield code="0">http://id.loc.gov/authorities/names/n2006039005</subfield>
  <subfield code="4">http://id.loc.gov/vocabulary/relators/trl</subfield>
</datafield>
```
URIs from BIBFRAME data

MARC data before conversion to BIBFRAME

<datafield tag="336" ind1="" ind2="">
    <subfield code="a">text</subfield>
    <subfield code="2">rdacontent</subfield>
</datafield>

MARC data transformed from BIBFRAME

<datafield tag="336" ind1="" ind2="">
    <subfield code="a">text</subfield>
    <subfield code="b">txt</subfield>
    <subfield code="0">http://id.loc.gov/vocabulary/contentTypes/txt</subfield>
    <subfield code="2">rdacontent</subfield>
</datafield>
URIs from BIBFRAME data

But URIs at the subfield level a problems for MARC – dependent on order

MARC data before conversion to BIBFRAME

<datafield tag="650" ind1=" " ind2="0">
  <subfield code="a">Women lawyers</subfield>
  <subfield code="v">Fiction.</subfield>
</datafield>

MARC data transformed from BIBFRAME

<datafield tag="650" ind1=" " ind2="0">
  <subfield code="0">http://id.loc.gov/authorities/subjects/sh2008113422</subfield>
  <subfield code="a">Women lawyers</subfield>
  <subfield code="0">http://id.loc.gov/authorities/subjects/sh85147640</subfield>
  <subfield code="v">Fiction</subfield>
  <subfield code="0">http://id.loc.gov/authorities/subjects/sh99001562</subfield>
  <subfield code="2">lcsh</subfield>
</datafield>
Punctuation

No punctuation at ends of elements in BIBFRAME

MARC data transformed from BIBFRAME

```
<datafield tag="300" ind1=" " ind2=" ">
  <subfield code="a">278 pages</subfield>
  <subfield code="c">21 cm</subfield>
</datafield>

<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Kissed by the rain</subfield>
  <subfield code="c">Claudia Winter ; translated by Maria Poglitsch Bauer</subfield>
</datafield>
```

Impact on MARC – no punctuation at ends of subfields
Summary

BIBFRAME to MARC data so far:
- MARC records are structurally sound
- URIs are carried into MARC
- No punctuation at subfield boundaries
- Provision activities in MARC 264’s (Production, Publication, Distribution, Manufacture, and Copyright Notice) not 260 (Publication, Distribution, etc. (Imprint))

Issues to reconcile
- MARC Bibliographic works
- URIs in subfields
Anonymous Resources, Blank Nodes, And Providers, Oh My!

Kevin Ford
What are Anonymous Resources?

First and foremost, they are RDF Resources like any other

Term/phrase "Blank nodes" frequently used as a synonym.
Anonymous Resources

<bf:Work rdf:about="http://bibframe.ex.org/18957985#Work">
  <bf:subject>
    <bf:Topic rdf:about="http://id.loc.gov/authorities/subjects/sh2008111698">
      <madsrdf:authoritativeLabel>Sisters--Drama.</madsrdf:authoritativeLabel>
    </bf:Topic>
  </bf:subject>
</bf:Work>

<bf:Work rdf:about="http://bibframe.ex.org/18957985#Work">
  <bf:subject>
    <bf:Topic>
      <madsrdf:authoritativeLabel>Sisters--Drama.</madsrdf:authoritativeLabel>
    </bf:Topic>
  </bf:subject>
</bf:Work>
Anonymous Resources

<http://bibframe.ex.org/18957985#Work> rdf:type
<http://id.loc.gov/ontologies/bibframe/Work> .
<http://bibframe.ex.org/18957985#Work> bf:subject
<http://id.loc.gov/authorities/subjects/sh2008111698> .

<http://id.loc.gov/authorities/subjects/sh2008111698> rdf:type
<http://id.loc.gov/ontologies/bibframe/Topic> .

<http://bibframe.ex.org/18957985#Work> rdf:type
<http://id.loc.gov/ontologies/bibframe/Work> .
<http://bibframe.ex.org/18957985#Work> bf:subject
_:Ne3ae2fe83bf743039944fd87abb6d7c9 .
_:Ne3ae2fe83bf743039944fd87abb6d7c9 madsrdf:authoritativeLabel "Sisters--Drama." .
_:Ne3ae2fe83bf743039944fd87abb6d7c9 rdf:type
<http://id.loc.gov/ontologies/bibframe/Topic> .
Anonymous Resources

RDF Resources may be identified by a URI or a "blank node".

Resources identified by a URI
   A (HTTP) URI is **shareable**
   A (HTTP) URI is **useable beyond a specific system**
   A (HTTP) URI is **re-usable within a specific system**

Resources identified by "blank nodes"
   A blank node is **not shareable***
   A blank node is **not useable beyond a specific system**
   A blank node is **not re-useable within a specific system**

* Technically it is not useable beyond the graph(s) in which the blank node exists but for the sake of this discussion...
Anonymous Resources

Get a bad rap but ...

They are a fact of life. Period.

Are part of the process.
Raw transformations frequently produce blank nodes,
i.e. Anonymous Resources
We do not (often or very much at all) have URIs in our data.

Should everything, really, have a URI?
URIs are commitments. You need to honor them. You need to manage them.
Bibframe and Blank Nodes

Although I am fairly zen about anonymous resources, they are my particular bugaboo at this moment.

Because:

1) I think it is important to give strong identity to those things that require strong identity

2) Using anonymous resources results in a tremendous amount of duplicate resources

3) Number of Anonymous resources are causing, or will cause, performance and scaling issues

4) Bibframe, like MARC, is about exchanging bibliographic information. Slimmer, more concise, more reusable resources are better for exchange.
Focus: Providers

<bf:Instance rdf:about="http://bibframe.example.org/11617736#Instance">
  <bf:provisionActivity>
    <bf:ProvisionActivity>
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication"/>
      <bf:place>
        <bf:Place>
          <rdfs:label>United States</rdfs:label>
        </bf:Place>
      </bf:place>
      <bf:agent>
        <bf:Agent>
          <rdfs:label>Columbia Pictures Home Entertainment</rdfs:label>
        </bf:Agent>
      </bf:agent>
      <bf:date>1982</bf:date>
    </bf:ProvisionActivity>
  </bf:provisionActivity>
</bf:Instance>
Focus: Providers

18M MARC Bib records = roughly 15M Anonymous Provider Agent resources

Many (most?) of these many represent the same entity
Consider such Providers as
  - Harcourt Brace
  - Penguin Books
  - Harper Collins

To wit:
  We have 1.2M unique "Providers" in those 15M.
Focus: Providers

So, we created a Providers file...

It is experimental.
Focus: Providers

Library of Congress Providers

URI(s)
- http://id.loc.gov/bentities/providers

Instance Of
- MADS/RDF MADSScheme
- SKOS ConceptScheme

Scheme Members
- BFEntities Providers

Change Notes
- 2018-12-19: modified

Alternate Formats
- RDF/XML (MADS and SKOS)
- N-Triples (MADS and SKOS)
- JSON (MADS/RDF and SKOS/RDF)
- MADS - RDF/XML
- MADS - N-Triples
- MADS/RDF - JSON
- SKOS - RDF/XML
- SKOS - N-Triples
- SKOS - JSON
Focus: Providers

Harcourt, Brace and Company

URL(s)
› http://id.loc.gov/bfentities/providers/939a6ff1ce3ea49822765a8261053bc8

Instance Of
› MADS/RDF CorporateName
› MADS/RDF Authority
› SKOS Concept

Scheme Membership(s)
› BFEntities Providers

Collection Membership(s)
› BFEntities Providers
› BFEntities Publishers

Variants
› Harcourt, Brace and Co.
› Harcourt, Brace and co.
› Harcourt, Brace and company
› Harcourt, Brace & Co.
› Harcourt, Brace and Company
› Harcourt, Brace & company
› Pub. for the United States armed forces by Harcourt, Brace and company
› Pub. for the United States Armed forces institute by Harcourt, Brace and company
› Harcourt, Brace, and Company
› Harcourt, Brace, and Co.
› Harcourt Brace & Co.
› Harcourt.Brace and company

Additional Information
› http://id.loc.gov/bfentities/agents/939a6ff1ce3ea49822765a8261053bc8
Focus: Providers

From BFEntities Providers

 المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي،

 URL(s)
 › http://id.loc.gov/bfentities/providers/55c079bae36fa51bd5cdf7b3257cf4

 Instance Of
 › MADS/RDF CorporateName
 › MADS/RDF Authority
 › SKOS Concept

 Scheme Membership(s)
 › BFEntities Providers

 Collection Membership(s)
 › BFEntities Providers
 › BFEntities Publishers

 Variants
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي
 › المملكة العربية السعودية، وزارة التعليم العالي، جامعة الإمام محمد بن سعود الإسلامية، عمادة البحث العلمي

 Additional Information
 › http://id.loc.gov/bfentities/agents/55c079bae36fa51bd5cdf7b3257cf4
Focus: Providers

Publication Test

Place: Place Test
Name: Name Test
Date: Date
Provision activity statement: Provision activity statement

Save changes
Focus: Providers

- **Name Test**
  - **Name**
    - harpercollins Children's (676184567f4414abb3f3d0262791edeb)
- **providers**
  - HarperCollins Children's (676184567f4414abb3f3d0262791edeb)
  - HarperCollins Children's Books (f42d333fbf121f5ccce16b22a7ae4135)
  - HarperCollins College Publications (14e1c9047447eaeca326f8f389c7efaa)
  - HarperCollins Interactive (ceea942b556f143df98d5894e8cecac9)
  - HarperCollins Leadership (af1092ff13a4b76cec071474037ed1d2)
  - HarperCollins Publishers India, a joint venture with the India Today Group (90571061c2aa58cf9055e532742d0f65)
  - HarperCollins Publishers Limited (8d69d74f8a68bf325f36f103720afa33)
Focus: Providers

<bf:Instance rdf:about="http://id.loc.gov/resources/instances/e18667955404035606954973019766450006035">  
  <bf:provisionActivity>     
    <bf:Publication>        
      <bf:place>        
        <bf:Place rdf:about="http://id.loc.gov/vocabulary/countries/nyu">New York (State)</bf:Place>        
      </bf:Place>    
    </bf:date>2019</bf:date>    
    <bf:agent>    
      <bf:Agent rdf:about="http://id.loc.gov/bfentities/providers/591fda23e3ac5afdf92d67cb54609a74">Random House Books</bf:Agent>    
    </bf:agent>  
  </bf:Publication>  
</bf:provisionActivity>
Thank You

http://id.loc.gov/entities/providers
LC Linked Data Service
id.loc.gov

Paul Frank
Library of Congress Linked Data Service
id.loc.gov
### Refine your results

#### Scheme
- Name Authority
- Subject Headings
- Entities Providers

#### Type
- Authority
- Simple Type
- Name
- Conference Name
- Corporate Name
- Topic

#### Collection
- Names - General
- Names - Authorized
- LC Subject Headings - Authorized
- LC Subject Headings - General
- LC Subject Headings - Geo May Subdivide
- Entities Providers

#### Created Date
- 2010s
- 1980s

#### Modified Date
- 2010s
- 2000s

### Results: 1-5 of 5

<table>
<thead>
<tr>
<th>#</th>
<th>Label</th>
<th>Vocabulary</th>
<th>Concept</th>
<th>Subdivision</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linked data</td>
<td>LC Subject Headings (LCSH)</td>
<td>Topic</td>
<td>3</td>
<td>sh2013002090</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data, Linked, Linked open data, LOD (Linked data), Open linked data, Open data, Linked</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>VISUAL (Workshop) (2014 : Linköping, Sweden)</td>
<td>LC Name Authority File (LCNAF)</td>
<td>Conference Name</td>
<td>3</td>
<td>nb2015019389</td>
</tr>
<tr>
<td>3</td>
<td>Data Link</td>
<td>Entities Providers</td>
<td>Corporate Name</td>
<td>3</td>
<td>c5b97064b5952a471e9b12d4d599b1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data Link</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>RUMOUR Workshop (2015 : Sibiu, Romania)</td>
<td>LC Name Authority File (LCNAF)</td>
<td>Conference Name</td>
<td>3</td>
<td>nb2016010545</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Workshop on Social Media and the Web of Linked Data (2015 : Sibiu, Romania)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>LINK Resources Corporation</td>
<td>LC Name Authority File (LCNAF)</td>
<td>Corporate Name</td>
<td>3</td>
<td>n88047251</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LINK (Firm) ; International Data Corporation ; LINK Resources</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do you need assistance with your search?
<table>
<thead>
<tr>
<th>Rank</th>
<th>Title</th>
<th>Authors</th>
<th>Year</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Wood, David. 1963. Linked data : structured data on the Web</td>
<td>Wood, David, 1963</td>
<td></td>
<td>(Work from Bib)</td>
</tr>
<tr>
<td>2</td>
<td>Jones, Ed. 1953. Linked data for cultural heritage</td>
<td>Jones, Ed, 1953</td>
<td></td>
<td>(Work from Bib)</td>
</tr>
<tr>
<td>3</td>
<td>Hartig, Olaf. Querying a web of linked data : foundations and query execution</td>
<td>Hartig, Olaf</td>
<td></td>
<td>(Work from Bib)</td>
</tr>
<tr>
<td>4</td>
<td>Linked data management</td>
<td></td>
<td></td>
<td>(Work from Bib)</td>
</tr>
<tr>
<td>5</td>
<td>Sikos, Leslie F.. Mastering structured data on the semantic web : from HTML5 microdata to linked open data</td>
<td>Sikos, Leslie F.,</td>
<td></td>
<td>(Work from Bib)</td>
</tr>
<tr>
<td>6</td>
<td>Hooland, Seth van. Linked data for libraries, archives and museums : how to clean, link and publish your metadata</td>
<td>Hooland, Seth van</td>
<td></td>
<td>(Work from Bib)</td>
</tr>
</tbody>
</table>

Future Discovery Layer?