Creating and Updating a BIBFRAME database

Library of Congress BIBFRAME Update

ALA Annual, New Orleans

June 24, 2018

bf: BIBFRAME

Moving from MARC to BIBFRAME at LC

- Revised BIBFRAME 2.0 data model and updated vocabulary
  - http://www.loc.gov/bibframe

- New MARC-to-BIBFRAME data conversion specifications and conversion programs
  - https://github.com/lcnetdev/marc2bibframe2

- Updated BIBFRAME record editor profiles

- Infrastructure improvements at the Library of Congress
  - Additional servers
  - Updates to database software and triple store (MarkLogic)

Creating the BIBFRAME database

- Entire LC MARC catalog converted to BIBFRAME, May 2017
  - 17+ million MARC bibliographic records converted to BIBFRAME Works, Instances, and Items
  - 1.2 million uniform title authority records converted to BIBFRAME Works
  - The BIBFRAME database is updated daily with new Work/Instance/Item descriptions from MARC bibliographic and authority records

- Current status
  - 19 million Works
  - 24 million Instances
  - 22.6 million Items
  - 4.3 billion triples

Searching the BIBFRAME database

Not designed to be public discovery system; limits and filters were created to facilitate analysis of the database

Match and Merge

- Work descriptions are created from MARC bibliographic records, MARC title and name-title authority records, and are created natively in the BIBFRAME editor

- Instance descriptions are created from MARC bibliographic records and in the BIBFRAME editor

- Software was developed to find matching Work descriptions, merge them, and link all Instances to the merged Work

Match and Merge

- Matching field creation:
  - MARC 130/240 uniform titles indexed as "nametitle"
  - Primary contributor (MARC 1XX) + title (MARC 245) fields are concatenated and indexed as "nametitle"
- Match
  - Find all the Instances (manifestations) for a Work
  - Determine if Work description already exists in file using the "nametitle" index
- Merge
  - Merge subjects and other Work information from Instances and add to Work description
  - Link Instances to Work description
- Repeat
  - Selective removal and reloading of descriptions will be necessary to merge Works and link Instances

Match and Merge

Work created from a MARC authority record with subjects added from merged bib record

Journal de physique. IV, Proceedings
Work from Authority (Untyped)

Work Journal de physique. IV, Proceedings (no98095786)
Title Journal de physique
Part number IV
Part title Proceedings

VariantTitle Journal de physique. Quatre, Proceedings
LCCN no 98095786
ISSN 1155-4339
Local identifier oca04767594
Source OCoLC

Topic Physics--Congresses. (11961785#Topic650-31)
Topic
MADS Auth label Physics

MADS Auth label Congresses

Topic Physics. (11961785#Topic650-32)
LCC Classification
Classification item number .J834
Classification number QC1

Match and Merge

Journal de physique
Instance (Untyped)

Instance Journal de physique. IV, Proceedings. (c0119617850005)

VariantTitle J. phys., IV
Source issnkey
Value issnkey

VariantTitle Journal de physique. IV
Title Journal de physique
Part number IV
Part title Proceedings

VariantTitle Journal de physique
Part number Quatre
Part title Proceedings

VariantTitle Journal de physique
Part number Four
Part title Proceedings

LCCN sn 98038158
ISSN 1155-4339
Source 7

Local identifier ocm39454418
Source OCoLC

Note Ceased publication
Physical details ill.
Note Title from cover.
issuance information The Proceedings have also consecutive numbering, 1-102, 1991-2002, cf. "List of colloquia" (includes earlier title) in v. 12, Pr11 (déc. 2002); with Feb. 2003, the consecutive numbering only is continued.
issuing body "Published under the scientific responsibility of the Société française de physique."
Provider statement Les Ulis, France : EDP Sciences, c1998-
Frequency completely irregular

Match and Merge

- However …

- Some types of works -- particularly music, audio and video -- need to be linked to other works but also stored as separate "RDA expression" works

- We may reload the MARC records, keep the works and link from the expression to the found work

- We are testing this process on translations

Homer. Iliad. Book 24
Work from Authority (Untyped)

Work	Iliad. Book 24 (nr2006000736)
Title	Iliad
Part number	Book 24
VariantTitle	Iliad, book XXIV
PrimaryContribution	
Person	Homer. (nr2006000736#Agent100-8)
Role	http://id.loc.gov/vocabulary/relators/ctb (ctb)
LCCN	nr2006000736
Topic (Agent)	Achilles (Mythological character) (425989#Agent600-21)
Topic	Trojan War--Poetry. (425989#Topic650-22)
Topic	
MADS Auth label	Trojan War
MADS Auth label	Poetry
Topic (Work)	Iliad. Book 24. (425989#Work600-23)
LCC Classification	
Source	DLC
Classification item number	.P24 1982
Classification number	PA4020
DDC Classification	
Source	DLC
Classification scheme edition	19
Classification scheme edition	full
Classification number	883/.01

Iliad, book XXIV

Instance (Untyped)

Instance	Iliad, book XXIV / (c0004259890002)
Title	Iliad, book XXIV
LCCN	81012208
ISBN	052124353X
ISBN	0521286204 (pbk.)
bibliography	Bibliography: p. 58-60.
Note	Includes index.
Supplementary material	Index present

Note	Publisher description
	http://www.loc.gov/catdir/description/cam022/81012208.html

Creative responsibility statement	edited by C.W. Macleod
Provider statement	Cambridge [Cambridgeshire] ; New York : Cambridge University Press, 1982.
Mode of issuance	http://id.loc.gov/vocabulary/issuance/mono (mono)

Publication

Date	(EDTF)1982
Place	http://id.loc.gov/vocabulary/countries/enk (enk)

Publication

Date	1982
Place	Cambridge [Cambridgeshire
Place	New York
Agent	**Cambridge University Press**

Extent	ix, 161 p.
Dimensions	20 cm.
Has Item	http://id.loc.gov/resources/items/c0004259890002

Instance Of
> Iliad. Book 24(nr2006000736)

Sibling(s)
> **Iliad, book XXIV / c0004259890002**
> c0004259890001

Has Items(s)
> c0004259890002

Homer. Iliad. Book 24

Work from Authority (Untyped)

Work	Iliad. Book 24 (nr2006000736)
Title	Iliad
Part number	Book 24
VariantTitle	Iliad, book XXIV
PrimaryContribution	
Person	Homer. (nr2006000736#Agent100-8)
Role	http://id.loc.gov/vocabulary/relators/ctb (ctb)
LCCN	nr2006000736
Topic (Agent)	Achilles (Mythological character) (425989#Agent600-21)
Topic	Trojan War--Poetry. (425989#Topic650-22)
Topic	
MADS Auth label	Trojan War
MADS Auth label	Poetry
Topic (Work)	Iliad. Book 24. (425989#Work600-23)
LCC Classification	
Source	DLC
Classification item number	.P24 1982
Classification number	PA4020
DDC Classification	
Source	DLC
Classification scheme edition	19
Classification scheme edition	full
Classification number	883/.01

Homer. Iliad. Book 24. English
Work from Authority (Untyped)

Work	Iliad. Book 24. English (n81085830)
Title	Iliad
Part number	Book 24

VariantTitle	Homer Iliad, book XXIV
PrimaryContribution	
Person	Homer. (n81085830#Agent100-8)
Role	http://id.loc.gov/vocabulary/relators/ctb (ctb)

LCCN	n 81085830
Language	English
Related resource	http://id.loc.gov/resources/works/nr2006000736
Related resource	http://id.loc.gov/resources/works/nr2006000736
Related resource	http://id.loc.gov/resources/works/nr2006000736
Related resource	http://id.loc.gov/resources/works/nr2006000736
Translation of	http://id.loc.gov/resources/works/nr2006000736

Expression/Translation Links
› Iliad. Book 24(nr2006000736)

Related Work(s), including stubs
› Iliad. Book 24(nr2006000736)

Match and Merge

1. Wayman, Eric. Tom Sawyer   **(Work from Authority)**
   Wayman, Eric.

1. Wayman, Eric. Tom Sawyer. Vocal score   **(Work from Bib)**
   Wayman, Eric.
   Ilford, Essex : Chappell, 1976.

Need work-to-work linking

Instance information (publisher) shouldn't stop works from merging

203. Dryden, John, 1631-1700. ... Dryden's Palamon and Arcite;   **(Work from Bib)**
   Dryden, John, 1631-1700.
   Boston, D. C. Heath & co., 1898.

204. Dryden, John, 1631-1700. ... Dryden's Palamon and Arcite;   **(Work from Bib)**
   Dryden, John, 1631-1700.
   New York, London [etc.] Longmans, Green, and co., 1897.

However …
Matching on titles isn't always perfect;
will need to add additional criteria

Untitled
Work from Authority (Untyped)

Work	Untitled (n42025799)	
Title	Untitled	
LCCN	n 42025799	
Local identifier	oca00024023	
Source	OCoLC	

Related resource (Work)

Work	Untitled (n42025799#Work410-10)	
Title	Untitled	
Contribution		
Organization	Friends of Photography. (n42025799#Agent410-10)	
Role	http://id.loc.gov/vocabulary/relators/ctb (ctb)	

Has Instance(s)
> [Untitled](c0198200010003)
> [Untitled](c0198300010003)
> [Untitled](c0199300010003)
> [Untitled](c0199600010003)
> [Untitled](c0200200010003)
> [Untitled](c0200542910003)
> [Untitled](c0200542980003)
> [Untitled](c0200543010003)
> [Untitled](c0200543040003)
> [Untitled](c0200543070003)
> [Untitled](c0200543130003)
> [Untitled](c0200543140003)
> [Untitled](c0200543150003)
> [Untitled](c0200543160003)
> [Untitled](c0200543180003)
> [Untitled](c0200543200003)
> [Untitled](c0200543210003)
> [Untitled](c0200543230003)
> [Untitled](c0200543260003)
> [Untitled](c0200543270003)
> [Untitled](c0200543450003)
> [Untitled](c0200543460003)
> [Untitled](c0200543470003)
> [Untitled](c0200543510003)
> [Untitled](c0200543520003)
> [Untitled](c0200543530003)
> [Untitled](c0200543540003)
> [Untitled](c0200543560003)
> [Untitled](c0200543570003)

… there are many more

BIBFRAME editor

Editor profiles customized by type of material

Bibframe Editor Workspace

Browse | Editor | Load Work | Load IBC

Create Resource

Monograph ▾

Notated Music ▾

Serial ▾

Cartographic ▾

Sound Recording: Audio CD ▾

Sound Recording: Audio CD-R ▾

Sound Recording: Analog ▾

Moving Image: BluRay DVD ▾

Moving Image: 35mm Feature Film ▾

Prints and Photographs ▾

Rare Materials ▾

Authorities ▾

BIBFRAME Work

Creator of Work (RDA 19.2)	Primary Contribution
Title Information	Work Title · Work Title Variation · Transliterated Title
Contribution (RDA 19.3 and 20.2)	Contribution
Subject of the Work (RDA Chapter 23)	Search subjects · Search subject components · Input subject strings
Form of Work	Form/Genre · RBMS term
Intended Audience (RDA 7.7)	Intended Audience (RDA 7.7)
Notes about the Work	Note
Contents (LC-PCC PS 25.1)	Contents note
Summary	Summary note
Classification numbers	Library of Congress Classification · Dewey Decimal Classification
Content Type (RDA 6.9)	Content Type (RDA 6.9)
	text 🗑
Language	Language
Illustrative Content	Illustrative content
Color Content (RDA 7.17)	Note
Supplementary Content	Supplementary Content
Related Works (RDA Chapter 25, Appendix J)	Related Work
Has BIBFRAME Instance	BIBFRAME Instance
Authorized Access Point Representing the Work (RDA 6.27.1)	Authorized Access Point Representing the Work (RDA 6.27.1) ＋

BIBFRAME Instance

BIBFRAME Instance

Instance Of
BIBFRAME Work

Title Information
| Instance Title | Variant Title | Parallel Title |

Chine | IGN Chine | IGN China
IGN 85102 Chine (2) | China

Statement of Responsibility Relating to Title Proper (RDA 2.4.2)
Statement of Responsibility Relating to Title Proper (RDA +
cartographie MairDumont
imprimée et diffusée en France par l'Institut national d...

Edition Statement (RDA 2.5)
Edition Statement (RDA 2.5) +
Édition 2-2013

Publication, Distribution, Manufacture Statements
| Publication Activity | Manufacture Activity |
| Distribution Activity |
MairDumont: Ostfildern

Copyright Date (RDA 2.11)
Copyright Date (RDA 2.11) +
© 2014

Series Statement
Series Statement
hasSeries

Mode of Issuance (RDA 2.13)
Mode of Issuance (RDA 2.13)
single unit

Identifier for the Manifestation
| ISBN | Other Identifier | Local system number |
9782758531579 | 2758531577

Notes about the Instance
Note
At head of title on panel: IGN
Insets: Beijing--I Hainan Dao

Media type (RDA 3.2)
Media type (RDA 3.2)
unmediated

Carrier Type (RDA 3.3)
Carrier Type (RDA 3.3)
sheet

Extent
Extent
1 map

Dimensions (RDA 3.5)
Dimensions (RDA 3.5) +
99 x 132 cm, folded to 25 x 12 cm

Base Material (RDA 3.6)
Base Material (RDA 3.6)
paper

Layout (RDA 3.11)
Layout (RDA 3.11)

Polarity (RDA 3.14)
Polarity (RDA 3.14)

Digital File Characteristic (RDA 3.19)
Digital characteristics

Uniform Resource Locator (RDA 4.6)
Uniform Resource Locator (RDA 4.6) +

Geographic Classification (MARC 052)
Geographic Classification
7820

Contributors (RDA 21.1)
Contribution

LC Control Number for the Manifestation (RDA 2.15)
LCCN
2018586394

Related Manifestation
Related Instance

Administrative Metadata for Instance
BIBFRAME 2.0 Admin Metadata
eng

Has Item
BIBFRAME Item
United States, Library of Congress

Server Connections

- The server hosting the BIBFRAME editor is linked to the server hosting the BIBFRAME database and http://id.loc.gov

- A cataloger can:
  - Extract any description for editing in the BIBFRAME editor
  - Update any description
  - Add additional Instances or Items
  - Link to Works and Instances already in the BIBFRAME database
  - Link to LCNAF, LCSH, LCGFT, LCMPT, MARC country codes, MARC language codes, and other standardized vocabularies at http://id.loc.gov
  - Send updates back to the BIBFRAME database and create new Works, Instances or Items

Open Issues

- The MARC-to-BIBFRAME conversion creates "stubs" for works in MARC 7xx tags; need a follow up process to unite these stub descriptions with full work descriptions

- Merging drops the 7xx headings from the work descriptions; illustrators, editors, etc., on subsequent editions are ignored

- Load sequence and system control numbers impact merging
  - http://...**e**2014431926 (work created in BIBFRAME editor)
  - http://...**c**018228499 (work created during MARC conversion)

Open Issues

- Need editing profiles for many types of workflows and materials; not fully defined yet

- Need ability to add a property/class "on the fly" while editing descriptions

- Descriptions that are retrieved from the BIBFRAME database, edited, and returned to the database need to be fully linked with existing descriptions

- Need the ability to "clone" a description – retrieve an existing Work or Instance, create a new description and save to the database with a new identifier

- Need to accept multiple data serialization schemes (XML, JSON, RDF)

What's Next?

- Continue to evaluate and adjust matching and merging in the BIBFRAME database and reload data as needed

- Ingest CIP and ONIX data

- Load Casalini RDF data

- Offer download of LC's BIBFRAME file for others to explore
  - Now available http://www.loc.gov/bibframe/implementation/

- Continue to improve editor

- Mapping from BIBFRAME to MARC

Thank you

Jodi Williamschen

Network Development & MARC Standards Office

Library of Congress

jowill@loc.gov

bf: BIBFRAME