Creating and Updating a BIBFRAME database

Library of Congress BIBFRAME Update
ALA Annual, New Orleans
June 24, 2018



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

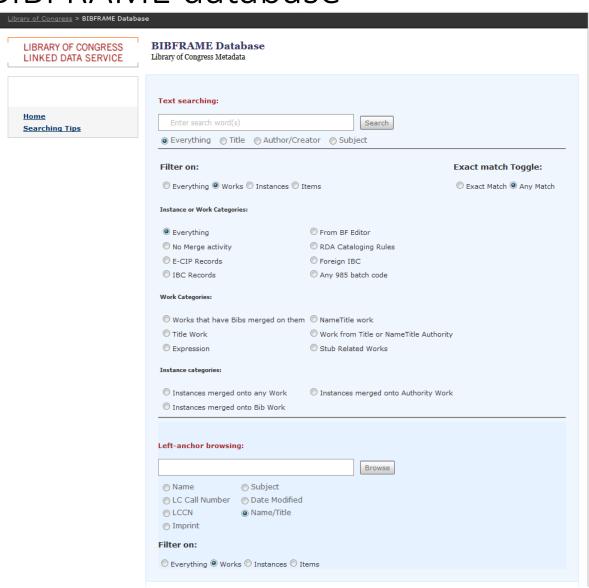
- Entire LC MARC catalog converted to BIBFRAME, May 2017
 - 17+ million MARC bibliographic records converted to BIBFRAME Works, Instances, and Items
 - 1.2 million uniform title authority records converted to BIBFRAME Works
 - The BIBFRAME database is updated daily with new Work/Instance/Item descriptions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



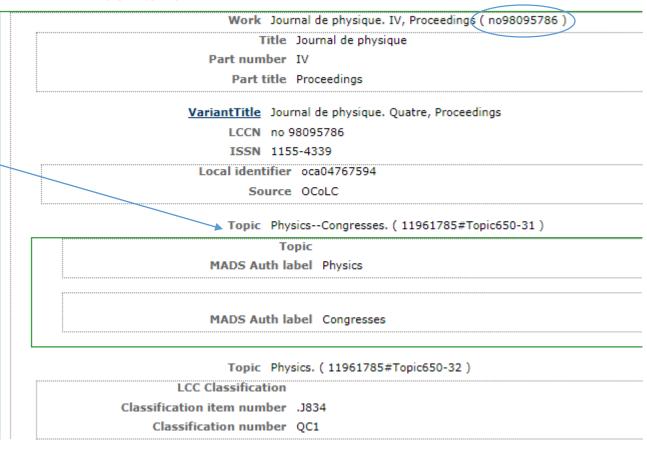
- Work descriptions are created from MARC bibliographic records,
 MARC title and name-title authority records, and are created natively in the BIBFRAME editor
- Instance descriptions are created from MARC bibliographic records and in the BIBFRAME editor
- Software was developed to find matching Work descriptions, merge them, and link all Instances to the merged Work

- 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

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

Journal de physique. IV, Proceedings

Work from Authority (Untyped)



		Instance Of	
Journal de physique Instance (Untyped)			> Journal de physique. IV, Proceedings(no98095786)
Thistorice (Onlyped)			
		Journal de physique. IV, Proceedings. (c0119617850005)	
Varian		itle J. phys., IV	
		Source issnkey	
		Value issnkey	
<u>v</u>	ariantTitle	Journal de physique. IV	
	Т	itle Journal de physique	
	Part num	ber IV	
	Part t	itle Proceedings	
	VariantT	itle Journal de physique	
	Part num	ber Quatre	
	Part t	itle Proceedings	
	VariantT	itle Journal de physique	
	Part num	ber Four	
	Part t	itle Proceedings	
	LCCN	sn 98038158	
	I	SSN 1155-4339	
	Sou	irce 7	
	Local ident	ifier ocm39454418	
		irce OCoLC	
	Note	Ceased publication	
Physi	ical details	ill.	
	Note	Title from cover.	
issuance in		The Proceedings have also consecutive numbering, 1-102, 1991-200 consecutive numbering only is continued.	2, cf. "List of colloquia" (includes earlier title) in v. 12, Pr11 (déc. 2002); with Feb. 2003, the
iss	suing body	"Published under the scientific responsibility of the Société française	de physique."
Provider	statement	Les Ulis, France : EDP Sciences, c1998-	
	Frequency	completely irregular	

- 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

Expression/Translation Links

- > Iliad. Book 24. English(n81085830)
 Related Work(s), including stubs
- <u>Iliad. Book 24. English</u>(n81085830)
 Has <u>Instance(s)</u>
- * Iliad, book XXIV (c0004259890002)
- > c0004259890001

Iliad, book XXIV

Instance (Untyped)

Instance Iliad, book XXIV / (c0004259890002) Title Iliad, book XXIV CAMBRIDGE GREEK AND LATEN CLASSICS. LCCN 81012208 ISBN 052124353X ISBN 0521286204 (pbk.) HOMER bibliography Bibliography: p. 58-60. Note Includes index. ILIAD Supplementary material Index present BOOK XXIV 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) EDITED BY C.W.MACLEOD Publication Date (EDTF)1982 Place http://id.loc.gov/vocabulary/countries/enk (enk) Publication Date 1982 Place Cambridge [Cambridgeshire Place New York

Extent ix, 161 p.

Dimensions 20 cm.

Has Item http://id.loc.gov/resources/items/c0004259890002

Agent Cambridge University Press

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

Expression/Translation Links

- Iliad. Book 24. English(#81085830)

 Related Work(s), including stubs
- > Iliad. Book 24. English(n81085830)
 Has Instance(s)
- > <u>Iliad, book XXIV /(c0004259890002)</u>
- > c0004259890001

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

> <u>Iliad. Book 24(nr2006000736)</u>

Related Work(s), including stubs

> Iliad. Book 24(nr2006000736)

1. Wayman, Eric. Tom Sawyer

Ilford, Essex: Chappell, 1976.

(Work from Authority)

Wayman, Eric.

Wayman, Eric.

1. Wayman, Eric. Tom Sawyer. Vocal score

(Work from Bib)

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)



Has Instance(s)

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

... there are many more

BIBFRAME editor

Bibframe Editor Workspace

Browse

Editor

Load Work

Load IBC

Editor profiles customized by type of material

Create Resource

Monograph ▼

Notated Music ▼

Serial ▼

Cartographic -

Sound Recording: Audio CD ▼

Sound Recording: Audio CD-R ▼

Sound Recording: Analog ▼

Moving Image: BluRay DVD▼

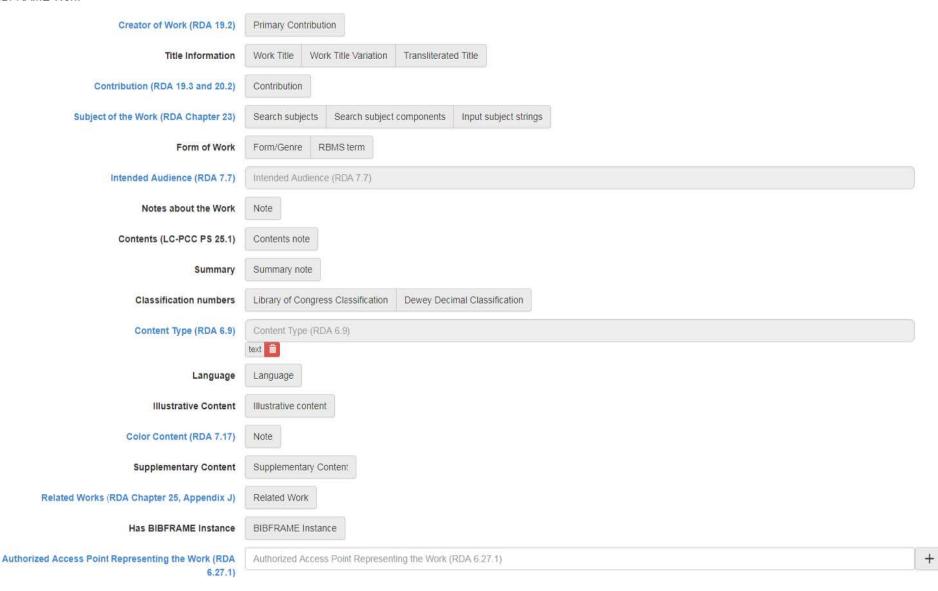
Moving Image: 35mm Feature Film ▼

Prints and Photographs -

Rare Materials -

Authorities -

BIBFRAME Work



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://...e2014431926 (work created in BIBFRAME editor)
 - http://...c018228499 (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

