

UC Office of the President

Presentation Grants

Title

From Creation to Reconciliation:Managing BIBFRAME Work and HubEntities at Scale—A Community Discussion

Permalink

<https://escholarship.org/uc/item/35j676bh>

Authors

Han, Myung-Ja K.

Heng, Greta

Lampron, Patricia

et al.

Publication Date

2026-07-11

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

From Creation to Reconciliation: Managing BIBFRAME Work and Hub Entities at Scale

– A Community Discussion

Myung-Ja (MJ) K. Han, University of Illinois Urbana-Champaign, mhan3@illinois.edu

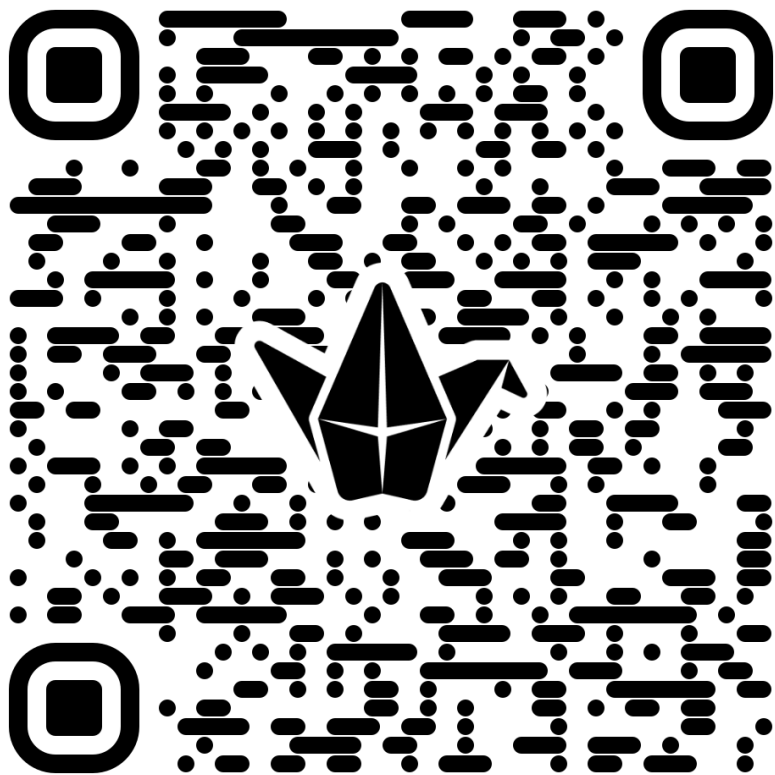
Greta Heng, San Diego State University, gheng@sdsu.edu

Patricia Lampron, University of California, Irvine, plampron@uci.edu

Deren Kudeki, iSchool, University of Illinois Urbana-Champaign, dkudeki@illinois.edu

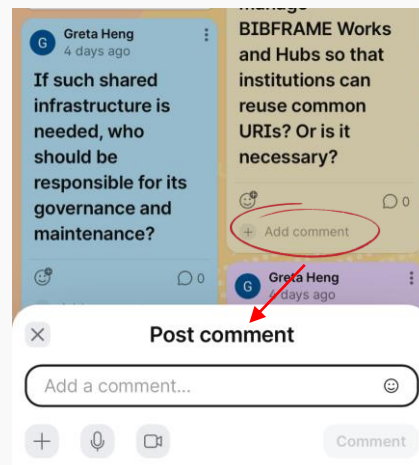
Motivations

- BIBFRAME has evolved to v3.0, introducing bf:Hub to support aggregation and collocation
 - Community efforts have focused primarily on **Creating BIBFRAME data**
 - **Limited attention to reconciling entities transformed from legacy MARC data**
 - MARC-to-BIBFRAME transformations typically generate Work and Instance
 - How reconciliation should be done
 - Where it should occur
 - How entities should be managed over time
-



We want to hear your thoughts!

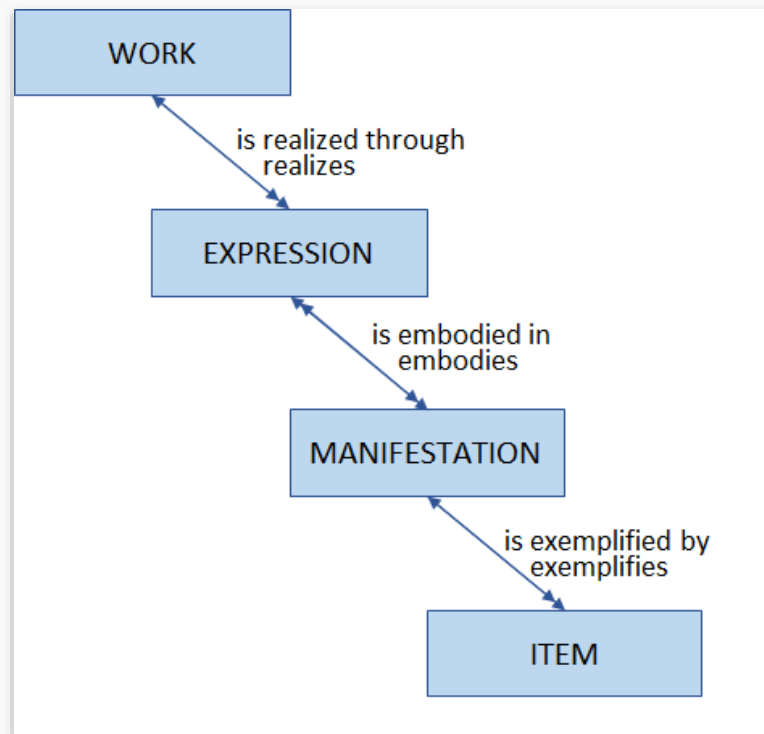
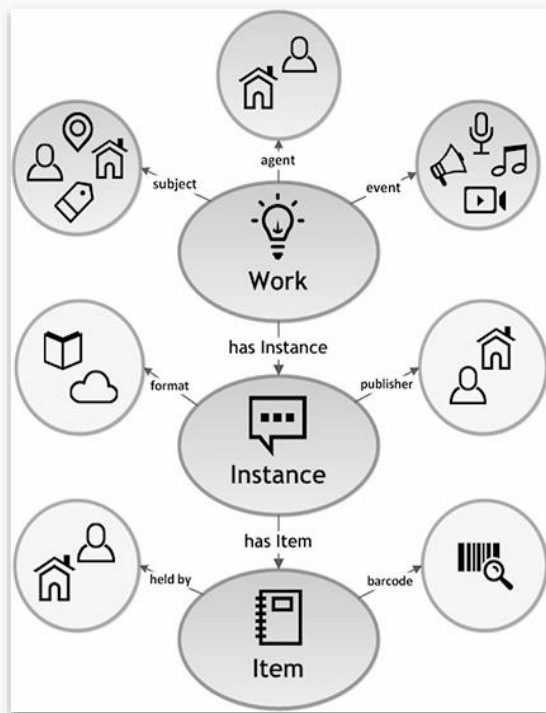
Scan the QR code and leave your comments



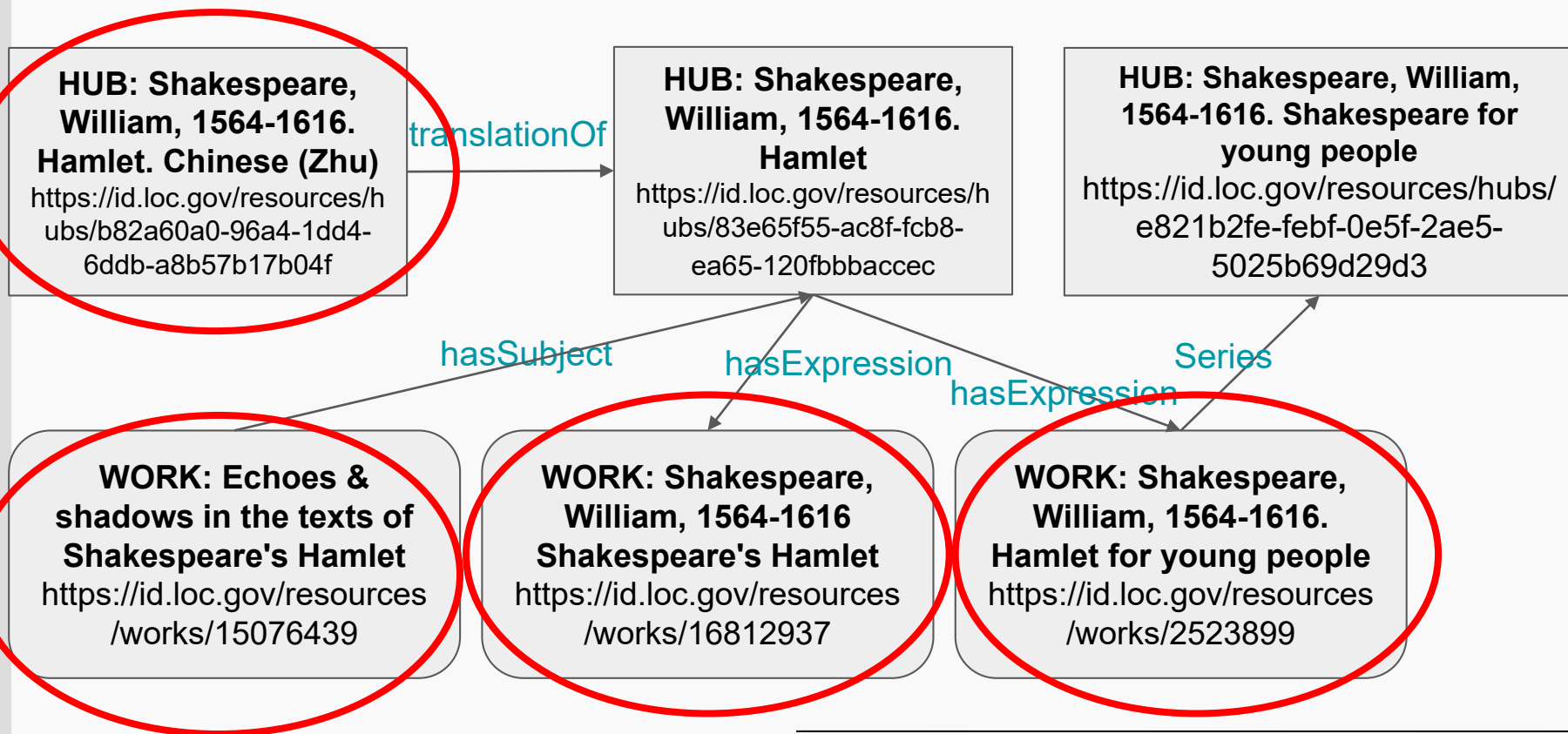
Outline

- BIBFRAME Data Model
- Our attempt: BIBFRAME Reconciliation
 - Methods
 - Findings
- Discussion

BIBFRAME Data Model



bf: Hub



BIBFRAME Reconciliation

- Hamlet: 8676 MARC records
 - Concerto: 86 MARC records
 - Random set of Locally digitized items: 236 MARC records
-
- Transformed to BIBFRAME using Library of Congress' ***marc2bibframe2***
-

Attempt 1&2&3

Attempt 4

Attempt 5&6

Attempt 7&8

Attempt 9&10

Run 1:

Exact text match on title.
Narrowest possible result.

Run 2:

Always choose the first
search result. Broadest
possible results.

Run 3:

Choose best search result
by fuzzy matching title
and notes. Both given
score from 0-1. Select
choice with highest sum
of scores.

Run 4:

Added:

- Fuzzy matching on contributors.
- Exact matching on languages.

Results:

All four dimensions being
equally weighted yields
obvious bad selections.
Title seems to be weighted
too heavily, while
contributor matches are not
impactful enough.

Run 5:

Try searching Wikidata
instead of LOC.
Results:
Very few Works found. No
way to find Hubs.

Run 6:

Fixed buggy cache, which
may have improved some
results.
Write results to BF.XML
for Virtuoso ingest. Add
new Work for each Hub
found.

Run 7:

Adjusted weights for:
Title, now 0-0.5
Contributors, now 0-2

Contributor can only
exceed 1 if matches are
found for multiple
contributors.

Did not adjust minimum
total score to qualify
Work as a match. This
makes multiple
contributor matches
heavily favored.

Run 8:

Bug fix: now query across
all variant titles.
Ignore ISNI contributor
links, so no Works throw
errors.

Run 9:

Use Uniform Title in Hub
search.

Contributor revisions:
Score on Primary & Secondary
contrib. separately. Sum
results for range of 0-3. Favors
matches on Primary contrib.

Increased minimum allowed
contributor score.

Scores are averaged across all
contrib. in records, rather than
on matches.

Run 10:

Bug fix: Search query now only
uses text from bf:Title
subfields, excluding
bf:lc:nonSortNum.

BIBFRAME Reconciliation - Findings

Attempt	UIUC Work	UIUC Hub	Concerto Work	Concerto Hub	Hamlet Work	Hamlet Hub	UIUC Work (Wikidata)	Concerto Work (Wikidata)	Hamlet Work (Wikidata)
1	11/295	0	17/97	0	1688/9837	0			
2	101/295	0	52/97	0	5484/9837	0			
3	71/295	0	47/97	0	3933/9837	0			
4	86/237	18/237	29/97	1/97	No data	No data			
5	80/237	17/237	29/86	1/86	No data	No data	0/237	1/86	162/8678
6	77/237	17/237	29/86	1/86	4319/8678	1128/8678			
7	64/237	10/237	23/86	0/86	3890/8678	772/8678			
8	65/237	12/237	No data	No data	No data	No data			
9	29/237	2/237	23/86	12/86	5010/8678	933/8678			
10	39/237	2/237	28/86	12/86	5992/8678	1319/8678			

Accuracy Evaluation

	UIUC Dataset	Concerto Dataset	Hamlet Dataset	Total
Match	8	5	2	15
Bad Match	2	4	3	9
Inconclusive	0	1	5	6
Total	10	10	10	30

Challenges - Duplicate Works?

Label	Dataset	Type	Subdivision	Identifier
1. Haughey, Joseph P. Teaching Hamlet in the twenty-first-century classroom	BIBFRAME Works	Work Text Monograph		23669756
2. Haughey, Joseph P. Teaching Hamlet in the twenty-first-century classroom	BIBFRAME Works	Work Text Monograph		23669899

Bibframe Instance

Title

Teaching Hamlet in the twenty-first-century classroom

Identified By

- Lccn: 2024018481
- Isbn: 9781475871821 (epub)
- Isbn: 9781475871807 (cloth) ([canceled or invalid](#))
- Isbn: 9781475871814 (paperback) ([canceled or invalid](#))

Note

- bibliography: Includes bibliographical references and index.
- description source: Description based on print version record and CIP data provided by publisher.

Extent

1 online resource

Provision Activity

Publication: [Maryland](#) 2024
Publication: Lanham: Rowman & Littlefield, 2024

Responsibility Statement

Joseph P. Haughey

Media

[computer](#)

Issuance

[single unit](#)

Carrier

[online resource](#)

Bibframe Instance

Title

Teaching Hamlet in the twenty-first-century classroom

Identified By

- Lccn: 2024018480
- Isbn: 9781475871807 (cloth)
- Isbn: 9781475871814 (paperback)
- Isbn: 9781475871821 (ebook) ([canceled or invalid](#))

Note

bibliography: Includes bibliographical references and index.

Extent

pages cm

Provision Activity

Publication: [Maryland](#) 2024
Publication: Lanham: Rowman & Littlefield, 2024

Responsibility Statement

Joseph P. Haughey

Media

[unmediated](#)

Issuance

[single unit](#)

Carrier

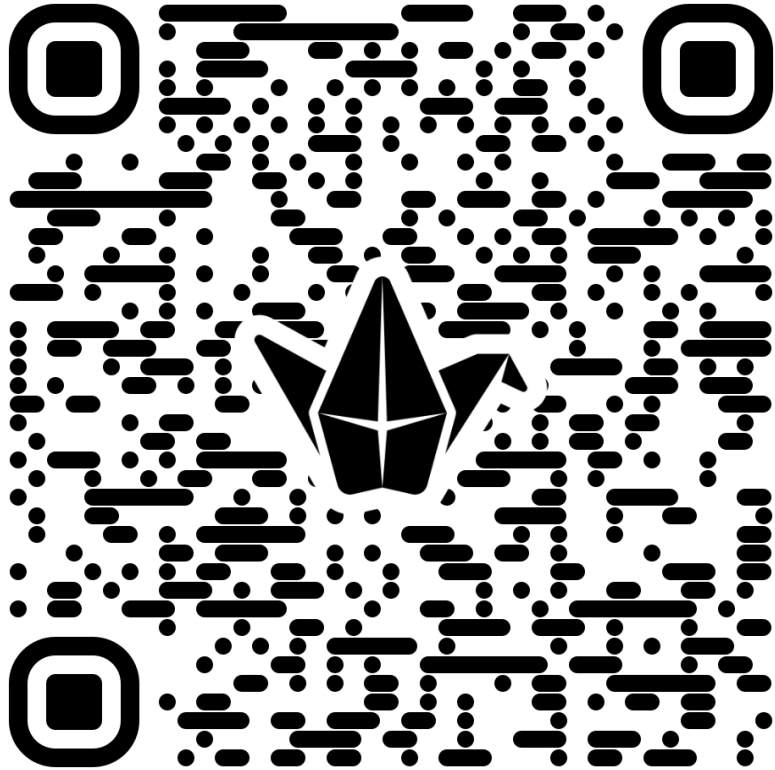
[volume](#)

Challenges - Data Characteristics

- Titles
- Contributors
- Notes

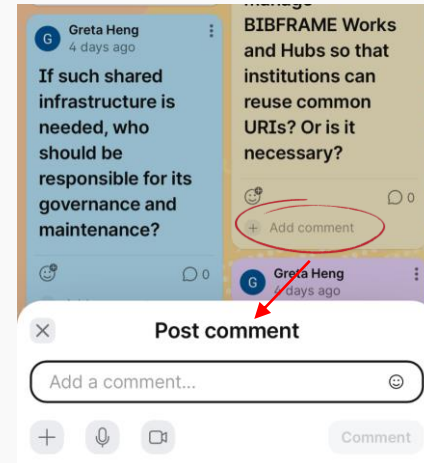
Heng, G., Kudeki, D., Lampron, P., & Han, M. J. (2026). Managing BIBFRAME Work and Hub Entities at Scale: Exploring Approaches to Large-Scale Reconciliation. *Cataloging & Classification Quarterly*, 1–21.

<https://doi.org/10.1080/01639374.2026.2655113>



We want to hear your thoughts!

Scan the QR code and leave your comments



Thanks!

Do you have any questions?

Credits: This presentation template was created by **Slidesgo**, and includes icons, infographics & images by **Freepik**

Acknowledgements:

Travel to the ELUNA 2026 conference for Patricia Lampron was supported by a LAUC Presentation Grant