

# Truly shared cataloging ecosystem development

SWIB Conference Workshop

2023-09-11 <https://bit.ly/SWIB23-ECO>

---

Facilitators: Steven Folsom, Jason Kovari, Simeon Warner  
(Cornell University Library)

# Introductions - Who are you?

- Name
- Institution
- Title/role
- What do you hope to get out of this workshop?

SWIB Code of Conduct: <https://swib.org/swib23/coc.html>

# Agenda

Introduction

Norms and Expectations

Facilitator presentation

Break

Facilitator presentation continued

Discussion: Benefits & Impediments

Breakout: Defining shared environment

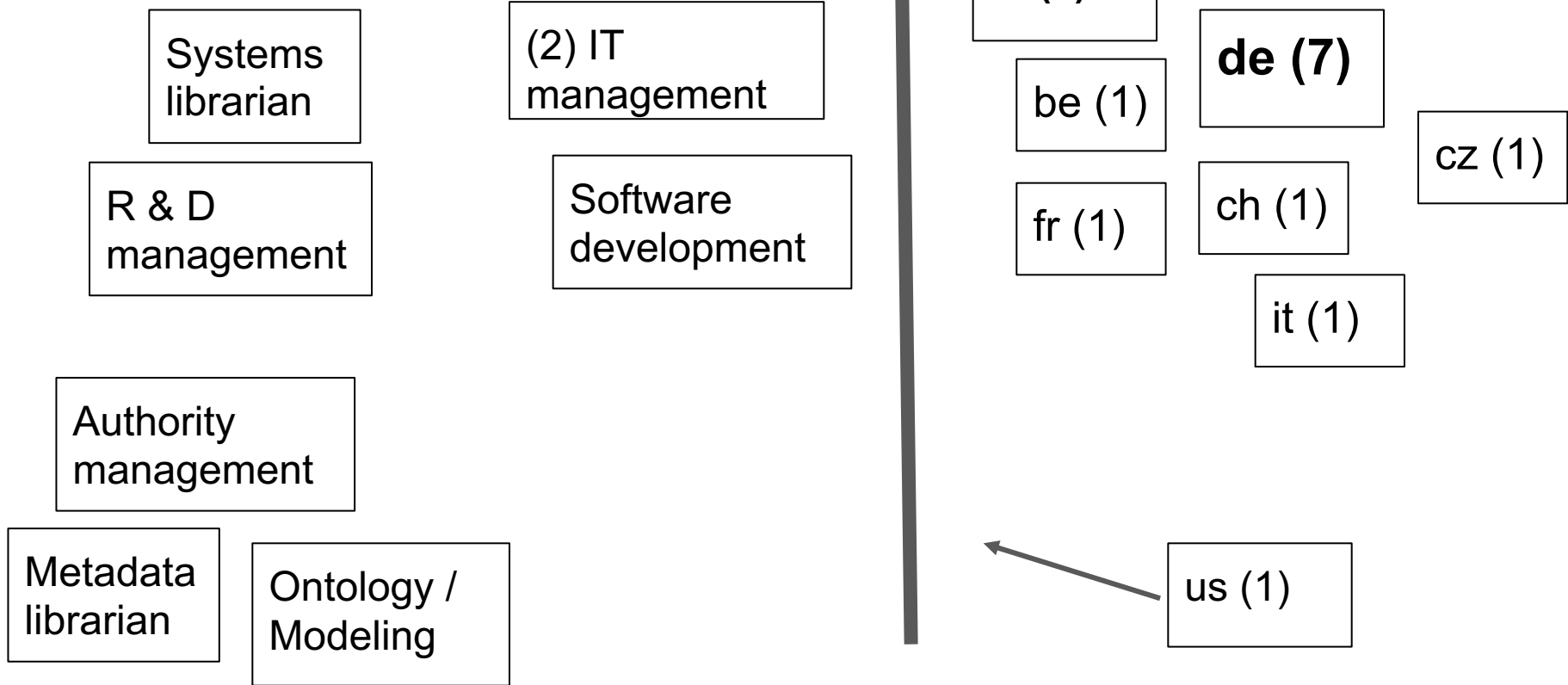
Break

Breakout: Boxes and Arrows

Discussion: Connections between shared environments

Conclusion

# What are your backgrounds?



# Where are we coming from?

- Jason, Steven, Simeon – We all work at Cornell University. We have experience of Cornell and other US university libraries
- Tom Cramer, Phil Schreur – Add experience from Stanford University and prior libraries
- Together, we have led a series of Andrew W. Mellon foundation grants exploring linked-data in libraries: the LD4L, LD4L Labs and LD4P grants spanning 2014-2023. I'll briefly describe some of the influence from the work in these grants
- Recent discussions have involved U Penn & Library of Congress
- We have also engaged with many others, including the SWIB and LD4 communities

We have benefitted from lots of discussion, but our perspective is **US centric** and also **large-academic-library centric**

# Experience from LD4L/LD4P Grants and related work

# Ontology Development

- LD4L Ontology, bibiotek-o
  - Concerns about early versions of BIBFRAME, desire to explore extensions
  - Work ended in 2017, focus on BIBFRAME feedback instead
- BIBFRAME extensions
  - E.g. ongoing maintenance of Art & Rare Materials (ARM) extension
- BIBFRAME very closely aligns with MARC – this is a mixed blessing:
  - Makes transition much easier and also reasonable fidelity ongoing conversion
  - Did not take full advantage of the opportunities of creating a new linked data format
  - “Weak” in terms of semantics/restrictions but inference etc. not a priority use case

```
-<rdf:RDF xml:base="http://bib.ld4l.org/ontology/">
  <owl:Ontology rdf:about="http://bib.ld4l.org/ontology/">
    <rdfs:label xml:lang="en-us">LD4L Proposed Bibliographic
    Ontology</rdfs:label>
    <dcterms:modified>2015-11-07</dcterms:modified>
    <dcterms:issued>2015-07-07</dcterms:issued>
    <skos:note xml:lang="en-us">PLEASE NOTE: This ontology
    LD4L 2014-2016 project
    refer to http://biblioframe.org/
    developed jointly by the
    projects.
  </skos:note>
  <rdf:comment xml:lang="en-us">This bibliographic ontology
  is an extension of the
  Sa
  to
  do
  BI
  th
  ma
  on
  re
  us
  da
  li
  an
  an
```

## bibliotek-o: A Framework for Modeling Bibliographic Metadata

### What is bibliotek-o?

bibliotek-o is an ontology framework for modeling bibliographic metadata based on the Library of Congress

### arm

## Version 1.0 of the Art and Rare Materials (ARM) BIBFRAME Ontology Extensions

NOTE: Version 1.0 replaces [version 0.1](#)

The ontology housed in this repository is being developed as an extension of the BIBFRAME ontology for generalized bibliographic description to provide specialized modeling in the archives, art and rare materials domains. This work was originally created under the LD4P 2016-2018 project. Work in this repository replaces work posted in the grant-funded LD4P repository: <https://github.com/LD4P>

Work on ARM version 1.0 was performed 2019-2021 under the auspices of the Art Libraries Society of North America's Cataloging Advisory Committee, the Association of College and Research Libraries' Rare Books and Manuscript Section's Bibliographic Standards Committee and the Society of American Archivists' Standards Committee.

### RDF files:

- Ontology: <https://w3id.org/arm/ontology/1.0/arm.rdf>
- Note Types vocabulary: [https://w3id.org/arm/vocabularies/1.0/note\\_types.rdf](https://w3id.org/arm/vocabularies/1.0/note_types.rdf)
- Origin vocabulary: <https://w3id.org/arm/vocabularies/1.0/origin.rdf>
- Physical Presentation vocabulary: [https://w3id.org/arm/vocabularies/1.0/physical\\_presentation.rdf](https://w3id.org/arm/vocabularies/1.0/physical_presentation.rdf)
- Relator vocabulary: <https://w3id.org/arm/vocabularies/1.0/relator.rdf>
- Status vocabulary: <https://w3id.org/arm/vocabularies/1.0/status.rdf>
- Typeface vocabulary: <https://w3id.org/arm/vocabularies/1.0/typeface.rdf>

# MARC → BIBFRAME Conversion

- Early grant work on converters to new ontologies
- Collaboration with Share-VDE on bulk-conversions for LD4P cohort and for PCC
- Exploration of Library of Congress converters

Data pools created by conversion have been valuable: understanding, source for copy cataloging, discovery work...

... BUT, less useful in helping develop environments and workflows for ongoing BIBFRAME data management. Need clear understanding of primary vs derivative records and synchronization rules

(Also, BIBFRAME → MARC converters exist)



# BIBFRAME RDF Editors

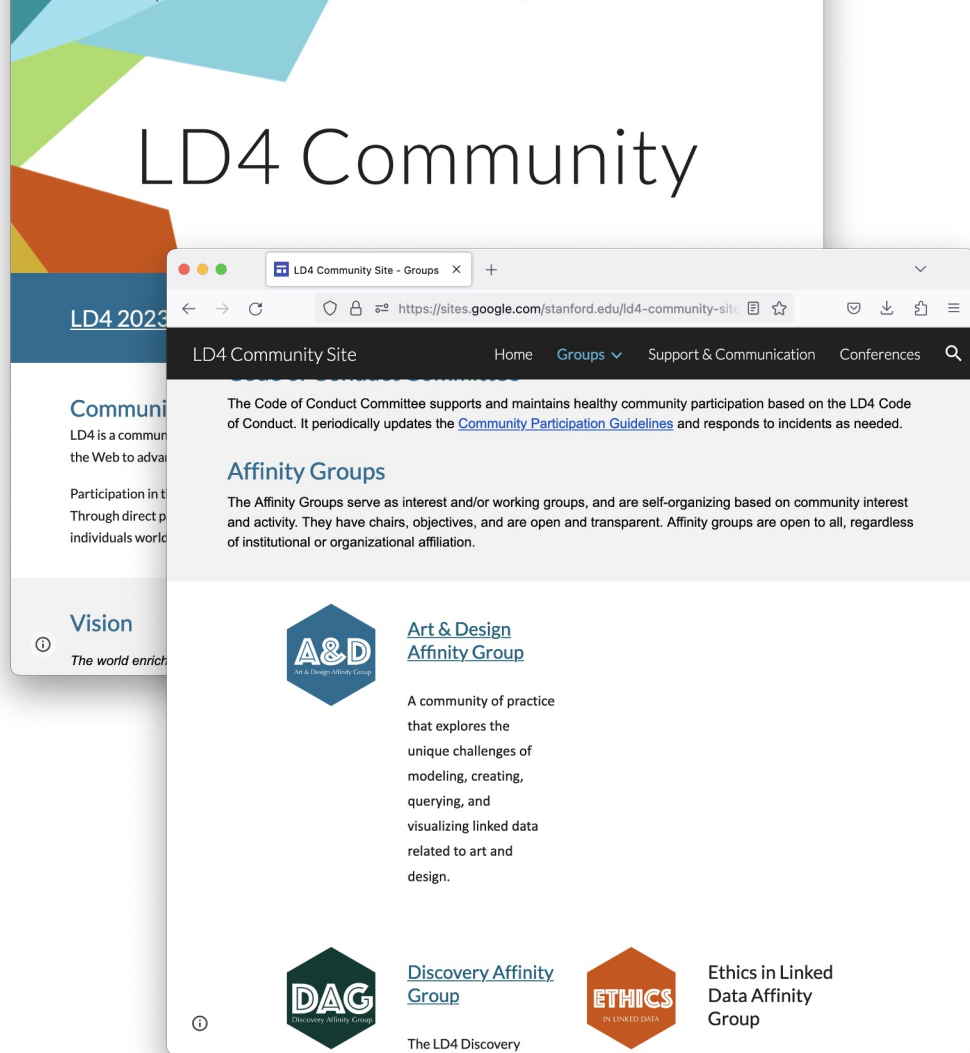
- Library of Congress
  - Developed [“BIBFRAME Editor” \(BFE\)](#), 2014–2021. Large scale tests
  - Rewritten as [“Marva”](#) starting in 2021
  - Being integrated with FOLIO by EBSCO for production use 2024 or 2025
- LD4L/LD4P
  - Early experiments with VitroLib (based in VIVO/Vitro editor) showed specialized facilities required for cataloging
  - Stanford team developed [“Sinopia”](#) editor, used by grant cohort, PCC and others

The screenshot displays the SINOPIA web interface. At the top, the header includes the 'SINOPIA' logo, navigation links for 'sfolsom', 'Help', and 'Logout', and a search bar containing the text 'Sinopia'. Below the header, a sidebar on the left lists navigation options: 'Dashboard', 'Editor', 'Resource Templates', and 'Actions'. The main content area is titled 'Monograph Work (BF2) Un-Nested' and features a 'WORK' button, a document icon, and an eye icon. A list of tabs on the left includes 'Contribution (Creator/Contributor)', 'Title Information', 'Form of Work', 'Date of Work', 'Place of Origin of the Work', 'In series', '(Geographic) Coverage of the Content', '(Time) Coverage of the Content', 'Intended Audience', 'Notes about the Work', and 'Bibliography'. The 'Contribution (Creator/Contributor)' tab is active, showing a form for editing a contribution. The form includes a 'Class' field with the value 'http://id.loc.gov/ontologies/bibframe/Work', a 'Contribution (Creator/Contributor)' section with a '+ Add another Contribution' link, and a 'Contributor' section with a '+ Add Name of Conference' link. The 'Contributor' section also displays the property 'http://id.loc.gov/ontologies/bibframe/agent' and the class 'http://id.loc.gov/ontologies/bibframe/Meeting'.



# Community

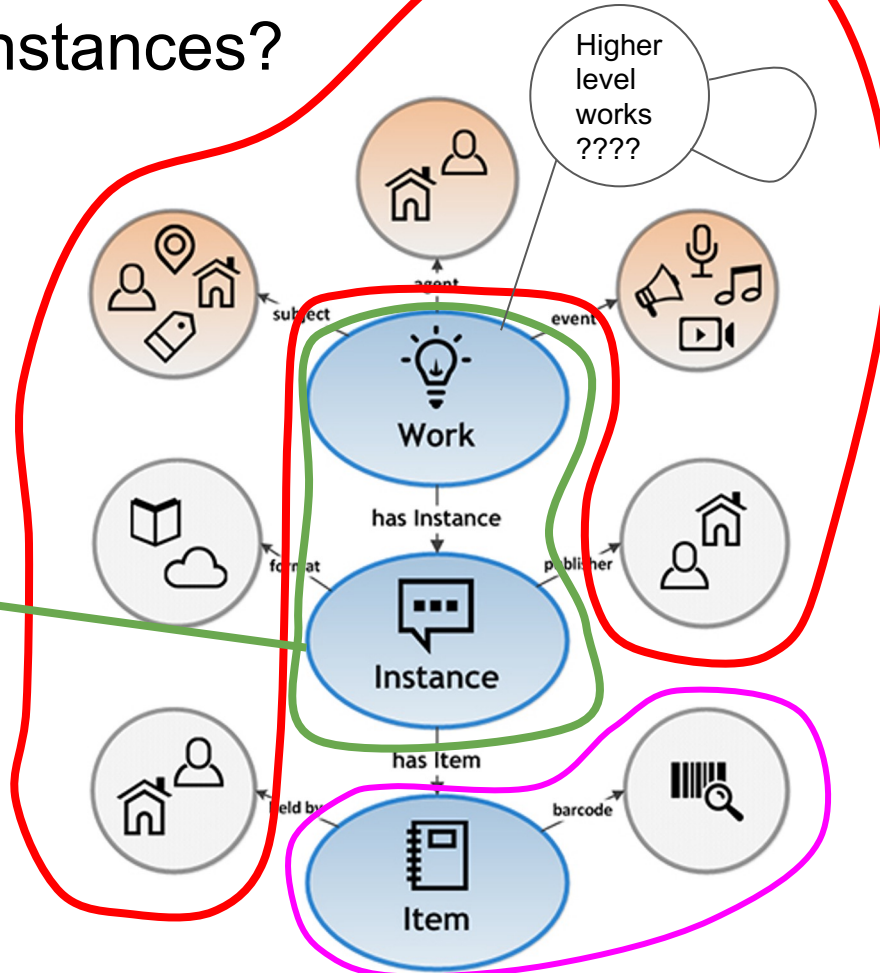
- SWIB & EuroBIBFRAME
- LD4 community:
  - Supported initially with help from LD4P grants
  - Affinity groups - Art & Design, Discovery, Ethics, Non-latin, Profiles, Rare Materials, Serials, Wikidata, Sinopia
  - LD4 Conference - Pandemic reinforced decision to have online not in-person LD4 conference - low cost, very broad participation



# Shared Works & Instances?

Much work to date  
copies these for  
every institution.

Real progress  
requires sharing.



Shared  
authorities,  
vocabularies &  
other linked  
entities (& a  
few strings)

Genuinely  
local data

# Two key motivations

## Improved discovery

- “Obvious” outcome of linked-data but needed practical demonstrations
- Experiments throughout the LD4L & LD4P grants, and by others
  - Specialized interfaces for particular data types or sub-domains
  - More general additions
  - User testing to verify utility / refine
- Improvements live in Cornell’s current discovery environment

→ Steven will share more details next

## Improved efficiency

- Not obvious with new technologies and more complex models
- Experiments have verified feasibility of components and workflows but have not been performed to get at efficiency
- We argue that proper use of linked entities, avoiding duplicative work, and avoid maintaining unnecessary copies can be more efficient

→ Jason will lead us into discussion of an efficient ecosystem

# Discovery: Lessons Learned

# Discovery: Background


- “Decorating” our MARC-based discovery environment
- MARC to RDF conversion pre-indexing
- “Linked Data” mostly serialized in RDF (but not limited to RDF)
- Cornell’s MARC often doesn’t include \$0’s or \$1s (yet)


# Discovery: Display


- Currently only displaying the linked data

## The wizard of Oz

"Illus. by Evelyn Copelman ; adapted from the famous pictures by W. W. Denslow."

Author, etc.:	<a href="#">Baum, L. Frank (Lyman Frank), 1856-1919.</a> <span>Author info »</span>
Format:	 Book
Language:	English.
Published:	New York, Grosset & Dunlap, [1956]
Subject:	<a href="#">Fairy tales.</a>
Description:	204 p., [10] leaves of plates. ill. (some col.) 22 cm.
Other contributor:	<a href="#">Copelman, Evelyn.</a> <a href="#">Denslow, W. W. (William Wallace), 1856-1915.</a>
Series:	<a href="#">Illustrated junior library.</a>
Notes:	first published in 1900 under title: The wonderful wizard of Oz.
Dedicated to:	Maude Gage Baum

Some of this information comes from [Wikidata](#)   
[Highlight the Wikidata data.](#)



### Availability

Olin Library [Hours](#)  
PZ8.B34 W8 1956 [Text](#)


✓ Available

[Request item](#)

[Request scan of article/chapter](#)

### Other forms of this work

[Wizard of Oz](#)

 Book English Online At the Library

[Wizard of Oz](#)



# Discovery: Caching/indexing to support local searching

- Gathering use cases for indexing until we can support updated caches
  - Advanced searching based on specific properties
  - Browsing based on related entities

[Back to item](#)

## Dynkin, E. B. (Evgeniĭ Borisovich), 1924-2014



Eugene Borisovich Dynkin (Russian: Евгений Борисович Дынкин; 11 May 1924 – 14 November 2014) was a Soviet and American mathematician. He made contributions to the fields of probability and algebra, especially semisimple Lie groups, Lie algebras, and Markov processes. The Dynkin diagram, the Dynkin system, and Dynkin's lemma are named after him. ([From DBPedia](#) [↗](#).)

Occupation:	Mathematician
Place of Birth:	Saint Petersburg (Russia)
Place of Death:	Ithaca (N.Y.)
Citizenship:	Soviet Union*
Educated at:	Moscow State University, MSU Faculty of Mechanics and Mathematics*
Affiliations:	Cornell University (starting 1977)

\* [From Wikidata](#) [↗](#)

Image: "[Eugene Dynkin 2003](#)" by Eugene Dynkin / Public domain

## Library Holdings

Total Works By: [64 Titles](#)

Total Works About: [2 Titles](#)

[Books \(61\)](#)

[Computer Files \(1\)](#)

[Manuscripts/Archives \(1\)](#)

[Videos \(1\)](#)

# Discovery: Attribution, Trust, and Data Quality

- Honoring licensing and supporting information literacy needs
- Making informed decisions about what to integrate
- Data quality
  - Ability to disallow/block data for certain entities and/or properties
  - Data improvements for data sources we now rely on

Notes: first published in 1900 under title: The wonderful wizard of Oz.

Dedicated to: Maude Gage Baum

Some of this information comes from [Wikidata](#) 

[Remove the Wikidata highlighting.](#)

## Other forms of this work

[Wizard of Oz](#)

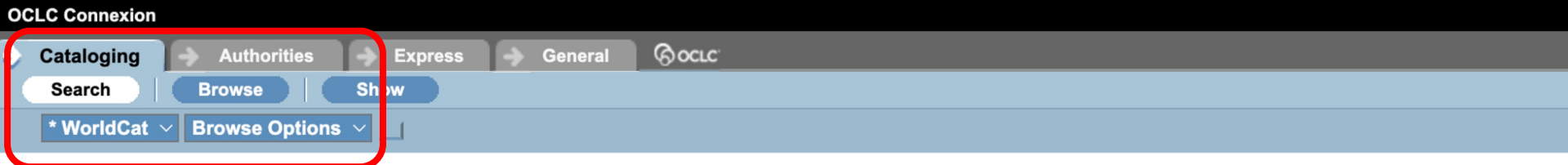
 Book English Online At the Library

[Wizard of Oz](#)

Linking to and Reusing RDF:  
Lessons Learned  
(Data Creation and Management)

# Libraries produce a lot of data, but rarely from scratch

- Deriving new descriptions from existing descriptions
  - Using similar works/publications/etc. to create similar descriptions for new works/publications/etc.
- Reuse and enhance existing descriptions
  - Finding an existing description and sometimes improving it, “copy cataloging”



**Welcome to the OCLC Connexion® service.**

We are discontinuing Connexion browser on 30 April 2024. Current Connexion browser users who prefer a web-based interface are encouraged to use the OCLC Discovery Manager today. Users of the Windows-based Connexion client can continue to use the locally installed Connexion client for their cataloging.

**Hello CTS!**

# Linking moments: Lookups when creating new data

Enter your own title and label

speech

Lookup with: LOC all genres (QA), MESH publication types (QA), GETTY\_AAT Objects\_\_Object\_Genres (QA)

Add "speech" as literal or select from:

LOC all genres (QA) (7) MESH publication types (QA) (0) GETTY\_AAT Objects\_\_Object\_Genres (QA) (8)

Speeches

Alternative Label: Televised speeches, Radio speeches, Academic addresses, Addresses (Speeches), Filmed speeches, Speechlets, Talks (Speeches), Public addresses, Legislative speeches, Orations

Citation note: (speech: A talk or formal address given to a public audience, extemporaneously or from a prepared text. In the 18th and 19th centuries, speeches were often separately published in print. Today, major public speeches are published in the New York Times and in the biweekly periodical Vital Speeches of the Day. Landmark speeches are usually anthologized.), (Legislative speeches), Commencement speech (A commencement speech or commencement address is a speech given to graduating students, generally at a university, generally in the US, although the term is also used for secondary education institutions) Baccalaureate service: (Baccalaureate addresses can range in length from under half an hour to as long as four hours) Inauguration (The "inaugural address" is a speech given during this ceremony which informs the people of his/her intentions as a leader) State of the Union (The State of the Union is the address presented by the President of the United States to a joint session of the United States Congress, typically delivered annually; State of the Union address) State of the State address (The State of the State Address is a speech customarily given once each year by the governors of each of the states of the United States (although the terminology for this speech differs for some states: in Iowa, the speech is called the Condition of the State Address; in Kentucky, Massachusetts, Pennsylvania, and Virginia it is called the State of the Commonwealth Address.) State of the City address (The State of the City Address - or State of the City Speech - is a speech customarily given once each year by the mayors of many cities in the United State, Canada, and the Philippines. Other international cities have also adopted the tradition), (speeches (documents). UF addresses (documents). Documents containing the text of any public address or talk. NT lectures; monologues; orations; sermons), (talk 1.2. An informal address or lecture), (Speeches Use Addresses; Addresses. UF Orations; Speeches. BT Discursive works. NT Academic addresses; Convicts' addresses; Lectures; Legislative addresses; Occasional addresses; Sermons), (talk: the act of talking formally about something before a group of people : a speech or lecture), p. 77 (speech: talk, conversation, discourse; statement, remark; formal address, oration; printed copy of this; speechlet: short speech)

Citation source: Reitz, J.M. ODLIS : online dictionary for library and information science, July31, 2014, Genre/form terms for law materials [report prepared by the Technical Services Special Interest Section of AALL], 2010; Wikipedia, July 31, 2014.; Art & architecture thesaurus online, July 31, 2014, Oxford dictionaries website, Sept. 7, 2016, Genre terms : a thesaurus for use in rare book and special collections cataloging, via WWW, July 31, 2014, Merriam-Webster dictionary online, Sept. 7, 2016, The 10 best graduation speeches of all time, via CNBC website, viewed July 31, 2014., Görlach, M. An alphabetical list of English text types, in Text types and the history of English, c2004:

Citation status: found

Broader: Discursive works

Narrower: Toasts (Speeches), Occasional speeches

Type: <http://www.loc.gov/mads/rdf/v1#Authority>, <http://www.loc.gov/mads/rdf/v1#GenreForm>, <http://www.w3.org/2004/02/skos/core#Concept>


Toasts (Speeches)

Citation note: (Toast (honor). A toast is a ritual in which a drink is taken as an expression of honor or goodwill. The term may be applied to the person or thing so honored, the drink taken, or the verbal expression accompanying the drink.), (toast: A call to a gathering of people to raise their glasses and drink together in honor of a person or thing, or an instance of drinking in this way), (toast: an occasion when people all drink together and say someone's name in order to express their admiration or their good wishes). (toast: a salutation or a few words of

# Linking moments: Reconciliation

- Reconciliation processes for strings
  - Adding identifiers to existing data, e.g. MARC \$0s
  - Converting to RDF
    - May include blank nodes or new URIs where there are no matches

## Reconcile column "keyword"

 » Access [Service API](#)

Reconcile each cell to an entity of one of these types:

- ☒ ULAN search  
/ulan
- ☐ TGN search  
/tgn
- ☐ AAT search  
/aat
- ☐ Search all Vocab  
/all

Also use relevant details from other columns:

Column	Include?	As Property
photo_id	<input type="checkbox"/>	<input type="text"/>
Photographer_username	<input type="checkbox"/>	<input type="text"/>
ai_service_1_confidence	<input type="checkbox"/>	<input type="text"/>
ai_service_2_confidence	<input type="checkbox"/>	<input type="text"/>
suggested_by_user	<input type="checkbox"/>	<input type="text"/>

# Linking moments: Supervised conversion (hybrid)

- Converting a single record into RDF
  - Some automated matching, some lookups
  - The cataloger is in control, overseeing the outcome
- E.g. [Oslo Public Library cataloging tool](#)

rett i ekstern kilde - Biblioteksentralen

Harry Potter og halvblodsprinsen  
2005 - Rowling, J.K.



---

Mangler hovedansvarlig: ☐

---



Hovedansvarlig:

**Aktør:**

Rowling, J.K.  

(1965- ) eng.

**Rolle:**

✕ Forfatter  


Legg til

---

Ikke eget verk: ☐

---

Søk etter eksisterende verk:

Harry Potter and the half-blood prince 

# Linking moments: At load time in applications

- As an application renders a page, a query is made to an API using a string match (implicit link) and brings back data.

## McBride, James, 1957-



American writer.

Occupation: Composer, Author, Journalist, Novelist, Musician  
Citizenship: United States of America\*  
Educated at: Columbia University, Oberlin College\*  
Affiliations: Columbia University, New York University, Oberlin College

\* [From Wikidata](#) 

Image: "[James mcbride 2013](#)" by [Larry D. Moore](#) / [CC BY 4.0](#)

### Library Holdings

Total Works By: [9 Titles](#)

Total Works About: [3 Titles](#)

 [Books \(8\)](#)

 [Videos \(2\)](#)

### Reference Information

Heading Type: Personal Name

Browse related items by call number:

[PS3613.C28](#)



# WOLFcon takeaways from the BIBFRAME in FOLIO panel

- With preliminary tools in place we need to define the ecosystem for BIBFRAME for cooperative cataloging
  - What the ecosystem looks like will likely differ depending on the region and libraries involved
  - Where is data coming from?
  - Where does it need to go so other libraries can find it?

Break  
10 minutes

# Shared Cataloging Ecosystem

# Scope

- Bibliographic data
- Not context entities, aka "authorities"

# Position of libraries in the ecosystem

- We are coming at this from a very US-centric view
- Cultural differences in ways that libraries organize

## MARC's legacy

- Copying records as historical practice
- Duplication allows for institutional-specific variation

## Linked Data development - if done wrong

- Reinforcement of old models of copying data
- Institutions unable to "give up" their local practices

# What do we feel is needed?

- Shared datastore
- Support for operations
- Cooperative decision making & ownership



# How do we shift the paradigm?

- Define & develop the components of a new ecosystem
- Drop idea that local bib data needs supercede shared requirements
- Build trust

# Benefits

Shared enhancements (3)

Avoid duplication (2)

Efficiency

Big player help smaller ones (2)

Improved quality in shared data (3)

Focus in essentials, not exceptions

Better collaboration

Services on shared data

Reduced complexity

Better quality assurance, data cleaning

No more MARC!

# Impediments

Trust (3)

Who can edit?

Letting go of control

Resourcing central services

Loss of prestige

Need shared cataloging rules & practices (3)

Data ownership

Complex problem resolution

Multilingualism

History

Including vendor metadata

# Breakout

Further define a shared environment based on the discussions including the right scale of this environment and what is the work that should happen in this shared environment?

## At what scale?

- Institutional (TOO SMALL)
- Group of institutions
- Consortium
- Regional (cf. some German networks, Flanders)
- National (cf. Finland)
- Multi-national
- Global (cf. wikidata)

Do institutions share one platform entirely (for item data) or just for the bibliographic data?

## Breakout - Defining a shared environment

Prompt questions:

- Where do you fall on the distributed vs. hub model?
- What operations need to be supported?
- Within a shared environment, what governance/rules are necessary?

Logistics:

- 30 minutes for discussion, 15 minutes to report out
- Select a notetaker who will report out
- Etherpad link: <https://etherpad.wikimedia.org/p/SWIB2023-TSC>

Break (Cake)  
30 minutes

## Breakout - Architecture: Boxes and Arrows

**What are the key components of a shared cataloging system and how does data flow between them?**

- What scale of collaboration do you imagine?
  - What components are shared vs distributed?
  - What other systems does it connect to?
- 
- If there are divergent views within the breakout group, feel free to draw multiple pictures
  - Feel free to use a drawing program of your choice, or draw on paper
  - Please be ready to talk through the picture(s) afterwards

# Discussion

- What are key similarities and differences between our pictures?
- How do different shared environments connect/collaborate?
- We assume most data will be open. Should and how could such systems support closed data too?