

# Linked data for opening up discovery avenues in library catalogs Huda Khan, Cornell University SWIB 2020

http://bit.ly/ld4pswib

#### Overview

- Acknowledgments and thanks
- Background and motivation
- Research and prototyping: Lightning round!
  - User feedback and lessons learned
- Ongoing and future work
  - Moving to production
- Community engagement



## Continuing LD4 Grants Evolution

- Linked Data For Pathways to Implementation (*LD4P2*)
  - "A collaborative project among four institutions (Cornell, Harvard, Stanford, and the University of Iowa) and the Program for Cooperative Cataloging (PCC)..."
- Linked Data For Production: Closing the Loop (LD4P3)
  - "LD4P3 aims to "close the loop" to create a working model of a complete cycle for library metadata creation, sharing, and reuse."

## Special Thanks and References

- Linked Data For Pathways to Implementation (LD4P2) 2018-2020
  - LD4P2 Discovery Wiki
  - Cornell Team: Steven Folsom, Jason Kovari, Lynette Rayle, John Skiles Skinner,
     Simeon Warner, Tim Worrall. DevOps: Greg Delisle, Robbie Douglas. (Special thanks to Frances Webb as consultant)
  - O Stanford Team: Chris Beer, Tom Cramer, Michelle Futornick, Jessie Keck, Jack Reed, Hilary Thorsen, Astrid Usong, Camille Villa, Jennifer Vine
  - O U of lowa: Dave Eichmann



<u>Linked Data For Production: Closing the Loop.</u> 2020-2022 (LD4P3)

## Special Thanks

- Greg Delisle (Cornell)
- Steven Folsom (Cornell)
- Michelle Futornick (Stanford)
- Jessie Keck (Stanford)
- Jason Kovari (Cornell)
- John Skiles Skinner (during LD4P2)
- Astrid Usong (Stanford)
  - Tim Worrall (Cornell)

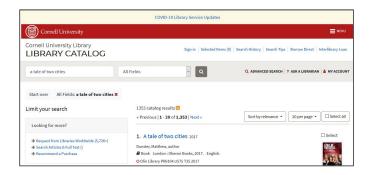


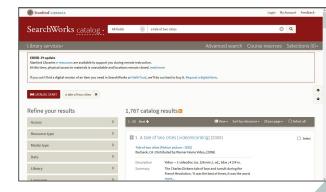


## **Background and Context**

### Discovery of library resources

- "Known item" search —>
   open-ended discovery
- Provide context around <u>entities</u>
- Include relevant relationships
- Link external data back to the catalog







## **Motivating Questions**

- Which <u>data connections</u> provide linkages or context?
- Which <u>discovery tasks</u> can we support using linked data?
- How can we support discovery of catalog resources through search engines?



## Wherefore Art Thou, Entities?





#### **External Data Sources**

Authority URI for author

http://id.loc.gov/authorities/names/n95046060



LCNAF Linked Data ~ 738 Million Triples



Lincoln at the Bardo



Work



Authority URI for subject

http://id.worldcat.org/fast/30184

Wikidata is highlighted but other data sources also included.







LCSH / FAST Linked Data LCSH: ~24.5 million triples



## Questions for You

- Additional linked data sources?
- Interactions and tasks to support?
- Production challenges and opportunities?
- Design ideas?
- User stories and scenarios?
  - E.g. BIBFRAME entities for discovery



# Lightning Round!

- LD4P2 focus: Design, develop, evaluate prototypes
  - Knowledge panels
  - Browsing
  - Semantic search/search suggestions
  - Schema.org output



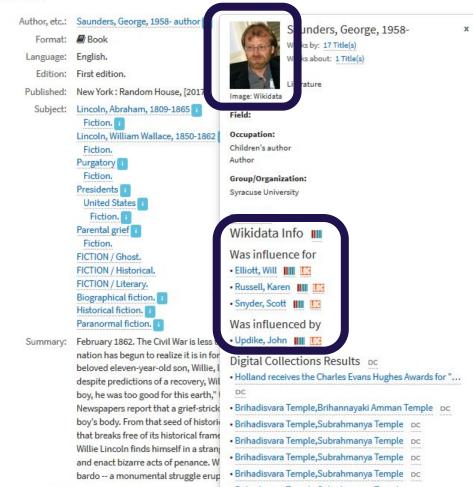
# Knowledge Panels and Context

Context around entities/related entities



a novel

George Saunders.

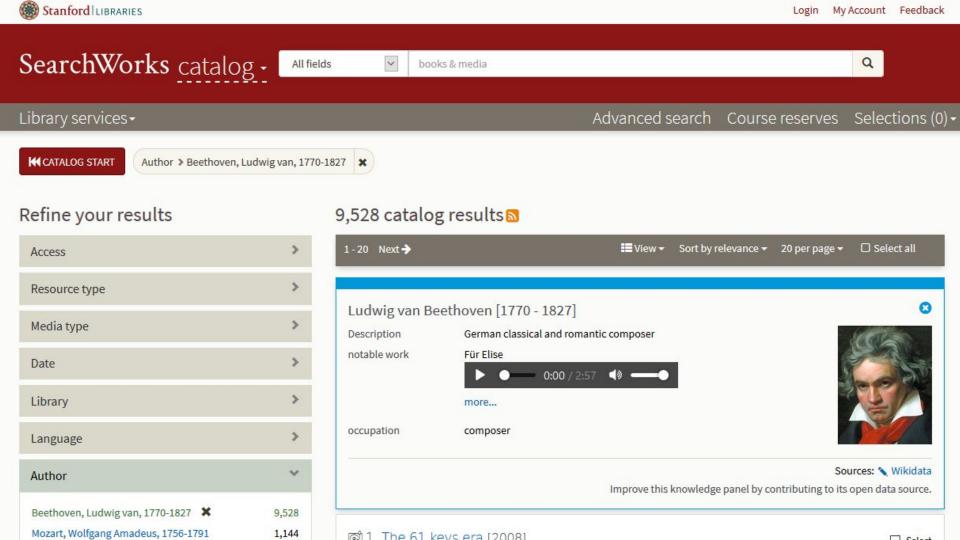


Availability

Olin Library Hours/Map
P53569.A7897 L56 2017 CText

O Checked out, due 10/25/19

Request item



# Knowledge Panels and Context

- Context around entities/related entities
- User feedback: "The Bomb Dot Com". Streamline information. Review properties used.
- Implementation: Client-side requests
- Data sources: Wikidata, Who's on First, LCNAF, LCSH,
   Discogs, digital collections



# Browsing

 Navigating categories and relationships while linking back to search



Search... 

All Fields 

Q ADVANCED SEARCH ? ASK A LIBRARIAN 

MY ACCOUNT

#### Limit your search

Access	~
At the Library	5,982,292
Online	2,200,304

Format	~
Book	7,044,037
Journal/Periodical	404,420
Microform	355,385
Musical Recording	235,901
Thesis	203,442
Video	106,232
Musical Score	103,059
Мар	65,092
Manuscript/Archive	17,429
Non-musical Recording	10,386
more »	

Language	>
Library Location	`

#### Welcome to the Cornell University Library Catalog.

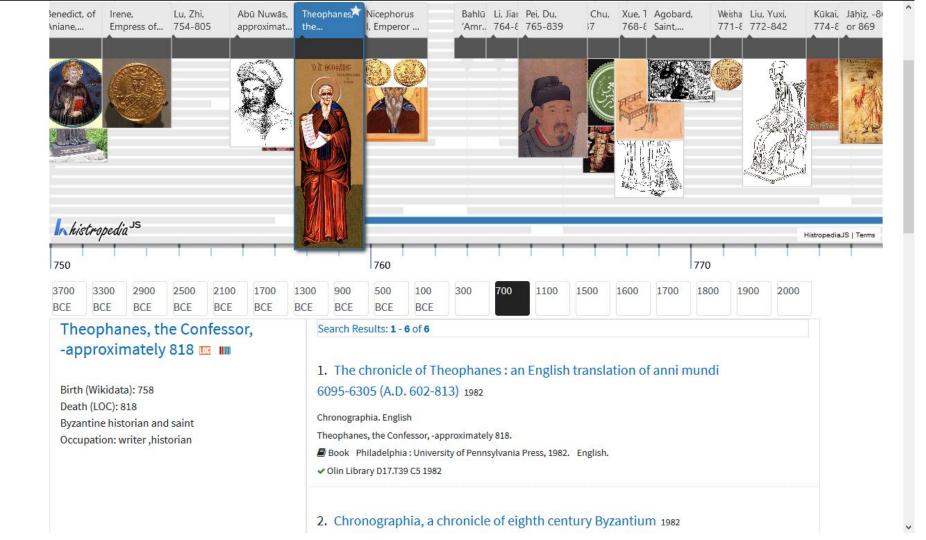
- · For help in searching the Catalog, try our Search Tips.
- For additional assistance, please Ask a Librarian.
- Or send us comments or suggestions that may help us improve the Catalog.

To search beyond Cornell holdings, follow the "Libraries Worldwide" link on the Catalog search results page, or use WorldCat.



#### OTHER RESOURCES

Articles & Full Text | Databases | E-journal Titles | Digital Collections



## Browsing

- Navigating categories and relationships while linking back to search
- User feedback: positive reactions for subject and timeline among novice users. Subject Matter Experts found less valuable.
- Implementation: Client-side and indices combining data sources



Data sources: LCSH, LCNAF, Wikidata, LCCN, Open Syllabus API

# Suggestions: Related Search Results

Provide <u>entity and search suggestions</u>









Book New Haven: Yale University Press, 2009. English.



4 Charles Dickons and



□ Select

☐ Select

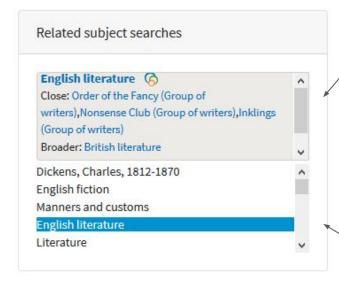


Mini knowledge panel for entity

### Entity suggestions:

- Annif Rest API
  - Training set: 194,555 entries matching full titles from the catalog with LCSH URI
  - User query -> LCSH suggestion





Mini knowledge panel for entity

### Entity suggestions:

- Facet strings for search results
  - FAST for subjects
  - LCNAF for agents



All Fields



Mini knowledge panel for entity

### Entity suggestions:

 Keyword search to Questioning Authority over LOC sources



## Suggestions: Related Search Results

- Provide <u>entity and search suggestions</u>
- User feedback: Makes catalog "stickier". Clarify how entities are related. Streamline.
- Implementation: Client-side requests and index for contemporary authors
- Data sources: Wikidata, LCSH, LCNAF, FAST, catalog metadata, Annif suggestions



# Suggestions: Autosuggest

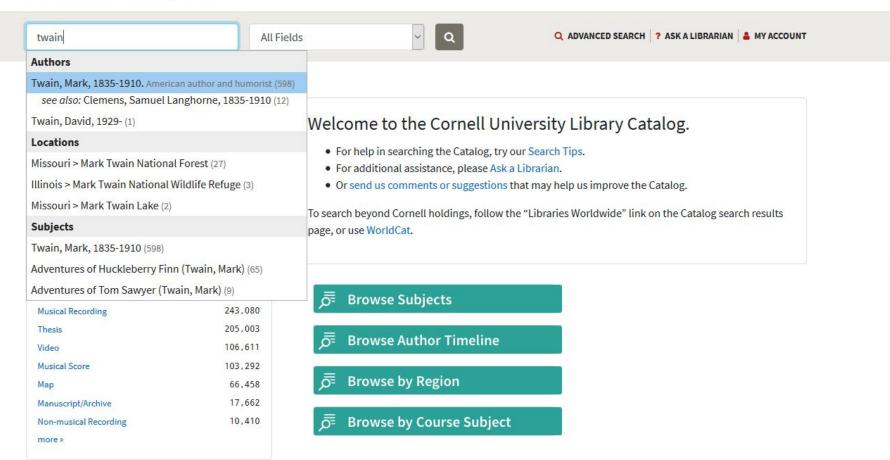
 Typeahead <u>suggestions for persons, locations, subjects,</u> <u>and genres</u> based on user query



#### Cornell University Library

#### LIBRARY CATALOG

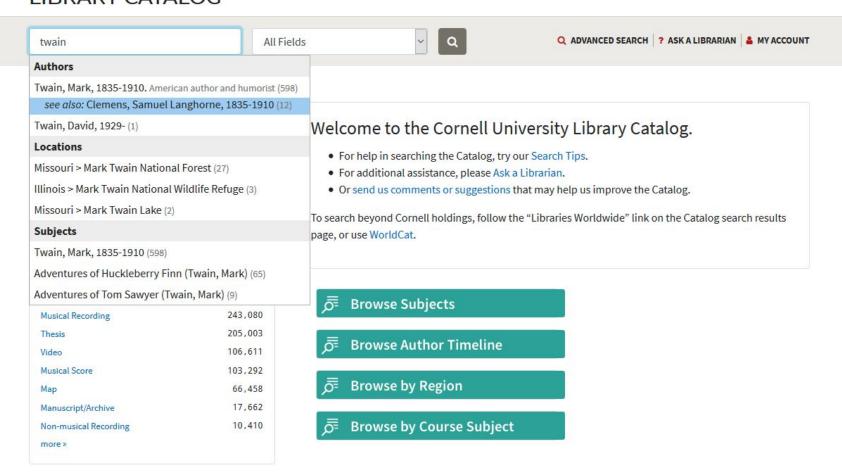
Sign in Selected Items (0) Search History Search Tips Borrow Direct Interlibrary Loan





### Cornell University Library LIBRARY CATALOG

Sign in Selected Items (0) Search History Search Tips Borrow Direct Interlibrary Loan



## Suggestions: Autosuggest

- Typeahead <u>suggestions for persons, locations, subjects,</u> <u>and genres</u> based on user query
- User feedback: Favorable and easy to understand.
   Clarification around shared pseudonyms. Misspelling support.
- Implementation: Client-side and index combining data sources



Data sources: catalog, LCNAF, FAST, Wikidata

#### **UI** Lessons

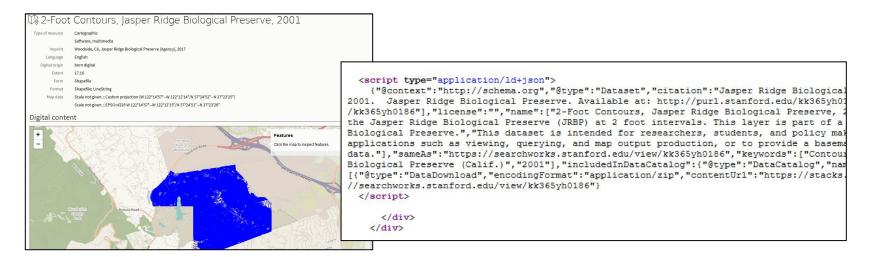
- Focused on non-specialists (undergrads/beginning researchers) to help widest range possible
- Subject matter experts had specialized workflows
  - They did not require as much assistance in viewing relationships between entities
- Interdisciplinary research means you may still be a novice in an area you need to research (See study on work discovery workflows)

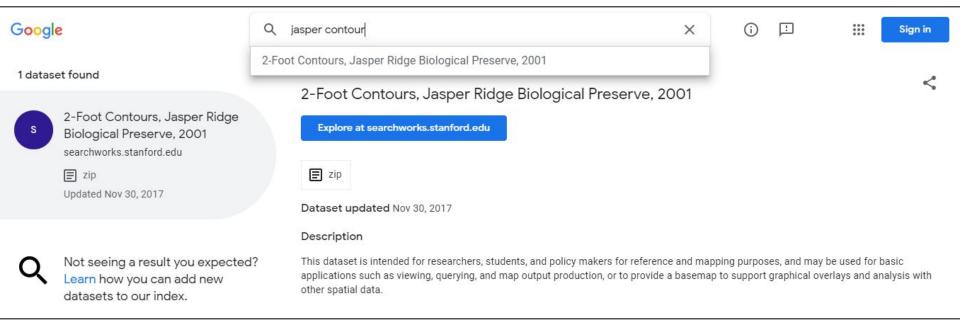
## Challenges and Opportunities

- Data: Consistency? Completeness? Ethical concerns?
  - UI and system-level handling?
- Client-side or index?
  - Performance, aggregation, search across combined data
  - Our How to handle updates?
- Additional work on accessibility





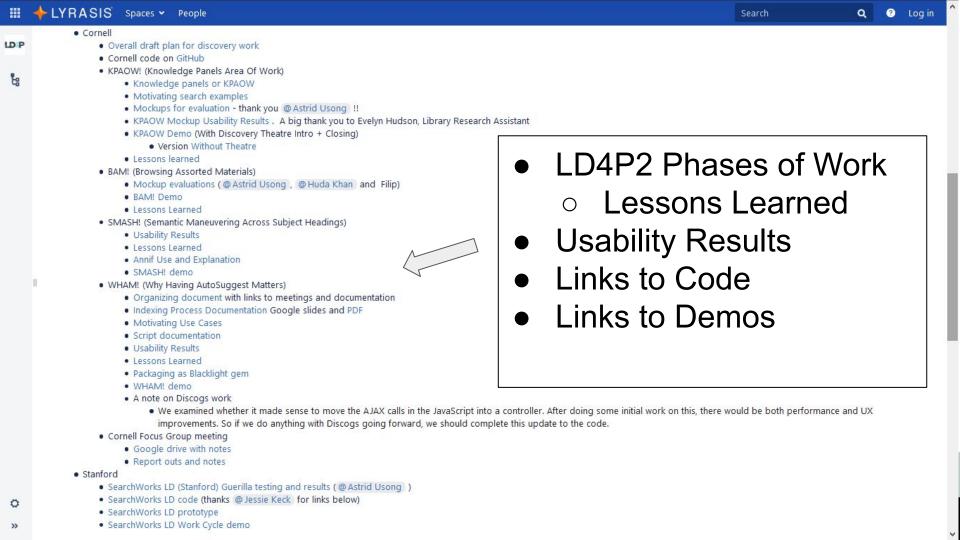




Search for keywords results in SearchWorks dataset being discovered.



Description and type of file appear match what is included in Schema.org information.



### Current Efforts: LD4P3

- Research and Design
  - New features or extensions of LD4P2 implementation
- Integrate with Production
  - Ideas from LD4P2 as well as LD4P3



# Current Efforts: Research and Design Strand

Research <u>user needs</u> and tasks

Identify data sources (library/external)

Identify connections with <u>library metadata</u>

Generate <u>design</u>

Develop <u>prototype</u>



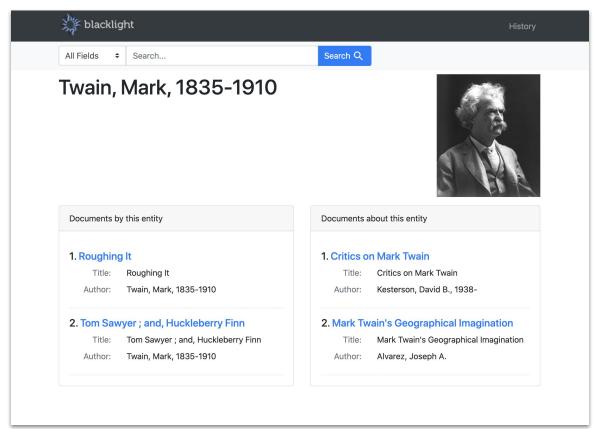
<u>Evaluate</u> prototype

### Current Efforts: LD4P3

- DASH!: Displaying Authorities Seamlessly Here!
  - Entity-specific dashboard
  - Cross-data source view of an entity within library catalog
    - E.g. Subject page
      - Broader/Narrower
      - PeriodO temporal and geographic information



# **URI Based Entity Pages**



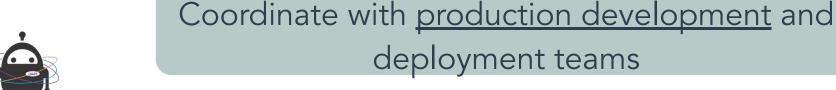


## **Current Efforts: Integrate Into Production**

Define data and architectural <u>dependencies</u>

Discuss and evaluate with production stakeholders

Refine and <u>"robustify"</u> as needed





#### High priestess of soul

Format: Musical Recording

Published: Philips [196-?]

Publisher number: Subject:

Popular music > United States.

Description: 1 disc. 33 1/3 rpm. microgroove. stereo. 12 in.

Contributors: Simone, Nina, 1933-2003, performer

PHS 600-219

Other

contributors:

Jack McMahon: Edited By [Tape Editor].

Del Shields: Liner Notes.

Hal Mooney: Producer, Arranged By.

Table of contents:

. Don't You Pay Them No Mind

• I'm Gonna Leave You

• Brown Eyed Handsome Man

• Keeper Of The Flame

• The Gal From Joe's

• Take Me To The Water

• I'm Going Back Home

• I Hold No Grudge

• Come Ye

• He Ain't Comin' Home No More

• Work Song

• I Love My Baby

Notes:

Popular songs, sung by Nina Simone.
Program notes by Del Shields on slipcase.



#### Availability

Music Library A/V (Non-Circulating) Hours/Map Rec 1630.18 S59 H6 OText

✓ Available On-site use

Request item

Request scan of article/chapter

Other forms of this work

High priestess of soul

♬ Musical Recording English Online

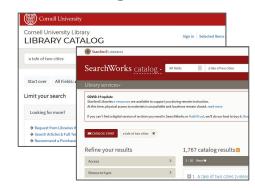
See all forms of this work



## Linked Data For Production: Closing the Loop







**Discovery Systems** 











**Linked Data Sources** 

Blacklight Code and Community

## Community Engagement and Review

- LD4 Conference -> Community
- Discovery Affinity Group
- Knowledge Panel White Paper
- Blacklight LD Working Meeting
- Wikidata Affinity Group
- Ethics in Linked Data Affinity Group



# And thanks to the SWIB community!



