KULA

Knowledge Creation, Dissemination, and Preservation Studies

Using Linked Data Sources to Enhance Catalog Discovery

Huda Khan, Claire DeMarco, Christine Fernsebner Eslao, Steven Folsom, Jason Kovari, Simeon Warner, Tim Worrall and Astrid Usong

See table of contents

Explore this journal

Cite this article

Article abstract

Our research explores how linked data sources and non-library metadata can support open-ended discovery of library resources. We also consider which experimental methods are best suited to the improvement of library catalog systems. We provide an overview of the questions driving our discovery experiments with linked data, a summary of some of our usability findings, as well as our design and implementation approach. In addition, we situate the discussion of our work within the larger framework of library cataloging and curation practices.

RESEARCH ARTICLE

Using Linked Data Sources to Enhance Catalog Discovery

Huda Khan
Cornell University

Claire DeMarco
Harvard University

Christine Fernsebner Eslao
Harvard University

Steven Folsom
Cornell University

Jason Kovari
Cornell University

Simeon Warner
Cornell University

Tim Worrall
Cornell University

Astrid Usong
Stanford University

Our research explores how linked data sources and non–library metadata can support open–ended discovery of library resources. We also consider which experimental methods are best suited to the improvement of library catalog systems. We provide an overview of the questions driving our discovery experiments with linked data, a summary of some of our usability findings, as well as our design and implementation approach. In addition, we situate the discussion of our work within the larger framework of library cataloging and curation practices.

Introduction

In recent years, there has been considerable work to transition from traditional library descriptive standards, notably MARC, to linked data approaches using Resource Description Framework (RDF)[1]-based description. These efforts range from grant-led research and development small-scale projects, such as Linked Data for Production's (LD4P) work on Bibliotek-o (Kovari, Folsom, and Younes 2017), as well as domain-specific efforts, such as the Art and Rare Materials BIBFRAME Ontology Extension (ARM 2021), to national-level production systems such as Sweden's LIBRIS (Kungliga bibliotek 2021) and the US Library of Congress (LOC) work on BIBFRAME (Library of Congress 2021). One of the claims that proponents of linked open data (LOD) make is

---

[1] RDF is a standard for representing semantic data on the web (W3C 2014).

that it will enable library data to be enriched through links to external LOD and that those data will in turn improve library discovery (Jin et al. 2016; Johnson and Estlund 2014; Naun 2019; Stahmer 2020).

In this paper, we will focus on a set of experiments that explore linked data's potential for improved user experience in discovery by providing additional context for entities described in the library catalog. Our experiments have been performed with a user-centered approach to designing, developing, and evaluating mock-ups and working prototypes, including using prototype additions to the catalogs at Stanford University and Cornell University. In this work, we incorporate information about entities in the catalog from RDF and non-RDF sources outside the catalog and show how we can use this information to provide more relevant links and resources in the catalog itself.

Both of the catalogs used in our experiments rely on traditional MARC metadata but nonetheless allow us to explore the results of using additional external data LOD sources in our discovery interfaces. The examples that follow do not require that the catalog data itself be modeled using an ontology such as BIBFRAME, though this work is relevant for metadata modeled using non-MARC paradigms that facilitate linking to external data sources.

Library Descriptive Standards Versus Open Data

User testing focuses on the systems built around library metadata but not on the usability affordances of the data itself, on whether gaps exist in the data, or on which data points are even necessary for our users. Inevitably, data usability for library patrons is tied to the systems through which those users experience that data. As we consider the inclusion of external data in our systems and look toward new data structures, however, it is even more important to assess both the systems and the data populating our systems.

Library descriptive standards for data have undergone a traceable evolution; MARC[2] was originally developed to represent the data that were within card catalogs, and BIBFRAME[3] well represents MARC data. Content standards such as AACR, AACR2, RDA, and RDA 3R, which dictate how we input data, have also developed along a direct line, with each iteration becoming more complex and offering the potential to support richer data. In practice, however, the evolution of these standards has not led to transformative changes in what data are captured.

Library discovery environments have evolved dramatically in the last few decades to include new features and more user-centered design; mimicking e-commerce websites, we have integrated faceted access, links for scanned copies of materials, tables of contents data, and cover images as part of individual records and search result pages (Golub 2018; Hall 2011; Wakimoto 2008; Wells 2020). Often these features use only data that the library community has carefully created and curated. Meanwhile, structured data created in non-library settings have increased in availability.

Library systems create and use controlled vocabularies to enable broad organizational structures. An organizational structure allows for reliable retrieval of known items and discovery of related items that reinforce the classification system itself. Many linked open data efforts, by contrast, seek to respond to the information that exists on the open web; these try to build connections between existing data points contained in that information to allow for discovery of new connections across and among sources without relying on a single or closed classification system (Snyder, Lorenzo, and Mak 2019).

There is a disconnect between descriptive practice and the expectations of many library users, including the struggle to retrofit subject headings into navigable facets and preferred headings that do not reflect our ethics or the terms commonly used in everyday language (McGrath 2007; Martin 2020). Linked open data initiatives, designed to take advantage of the connectedness of online information, reflect an attempt to meet online researcher expectations. Karen Coyle makes a strong argument that the data points catalogers are directed to prioritize do not always line up with known discovery needs (Coyle 2016); in turn, our MARC records and library linked datasets derived from MARC have struggled to respond to and account for user needs and preferences. Using this framing, linked open data beyond libraries (and library standards) offer the opportunity to foreground metadata usability and discovery interface design.

By and large, library discovery environments are bounded worlds. While libraries have discussed integrating non-library data in discovery environments for years, the community has not done so beyond a handful of examples like table of contents and cover art (Elguindi and Schmidt 2012). We cling to the notion that our data are more accurate, and consideration of integrating non-library data raises questions about the

---

[2] MARC is a data standard used for managing library resources in library management systems as well as user-focused discovery environments; first developed in the 1960s, MARC persists as the primary data model to support description and resource sharing across libraries (Library of Congress 2020).

[3] Developed and maintained exclusively by the LOC, BIBFRAME is an ontology designed as the successor to MARC; an OWL ontology, BIBFRAME is intended to support LOC's transition from MARC to linked data (Library of Congress n.d.).

unknowns and potential inaccuracies in those datasets. However, we regularly see inconsistent metadata and errors in library data (Martin and Mundle 2010; Wisser 2014; Yasser 2011). Further, vendor-provided discovery environments often lack the flexibility that would allow librarians to consider fully how to use external data.

Many current library discovery interfaces reflect an attempt to apply full-text keyword searching across a system designed for searching within particular MARC fields or browsing within metadata indices. The notion that an expertly crafted Boolean search will return all relevant resources for a given term presents a false sense of completeness and accuracy; further, this approach does not represent the search methods of most library users (even advanced researchers). This false perception is in large part due to limited resources that lead to varied levels of cataloging and a general lack of agreement on the value of certain elements in catalog records (Gross, Taylor, and Joudrey 2015).

Some within the library community have raised the potential inaccuracy or incompleteness of linked data as a possible challenge for integrating them in library systems (Schilling 2012). Although we recognize the need to present reliable data to users, our own catalog data are far from perfect. Requiring perfectly accurate and complete data linkages for the use of any linked data sources would prevent more widespread adoption of linked data initiatives. With its ability to model context around entities and to relate entities across different domains and data sources, the linked data framework aligns well with established user expectations for search: that a search box surfaces the most relevant results and provides contextual data to help users navigate to and discern between results (Dempsey 2012; Diao 2018). Put more simply, library users are internet searchers who look to external data of variable quality on the web to make choices about search results.

Knowledge panels are one well-established,[4] usable pattern for demonstrating connections between sources of information online. The broader implications, however, for incorporating knowledge panels into a library discovery system raise concerns about accuracy and completeness. By making the connections between sources of information more visible and accessible at a perceived point of need, we are making explicit a problematic reality: even the most strictly controlled information ecosystem may present inaccurate information and will always be incomplete.

We can support users' information-seeking behaviors by not only focusing on descriptive standards but also by including contextual subject matter, designing usable and appealing interfaces, and enhancing underlying software. Again, we see the need for a coalition of experts to push linked data programs ahead by demonstrating benefit for users across the library and information science communities at large. Rather than leading from the perspective of descriptive standards and vocabulary control, our experiments in linked data programs attempt to lead explicitly from a position of user needs, requirements, and preferences and to leverage information that is siloed in other environments.

Centering UX Design and Development

Beyond Known Item Search

In the course of discovery and information seeking, end-users can engage in multiple kinds of tasks. Wakeling et al. (2017, 2176) note that users conduct both "known item" searches, where users look for a particular item, and "unknown item" searches, which range from searches on a particular topic to searches to identify "unknown titles by a known author." Palagi et al. (2017, 5) characterize exploratory search as a search that is open-ended and during which a user has an evolving information need and search process, a "serendipitous attitude," and possibly carries out "several one-off pinpoint" searches. They note an overlap between exploratory search and Bates's (1989) "berrypicking" model of information seeking, according to which users dynamically change their search strategies as they discover new and/or different search terms from the results they are experiencing.

Known item search, or the ability to find a resource in the catalog given the title or author of a work, has long been considered an important feature. Library catalog search would benefit from supporting other more open-ended discovery paradigms such as browse and exploration (Trapido 2016). In their user studies with undergraduate and graduate students, Boettcher (2020) found that experienced researchers find browsing helpful as their fields become more interdisciplinary and they need to expand their knowledge to adjacent fields. These researchers also value serendipitous discovery while browsing. McKay et al. (2019, 1384) state that "given the importance of serendipity and browsing to information work, they are information-seeking strategies we lose at our peril." Gusenbauer and Haddaway (2021, 141) posit that discovery systems can help support exploratory search by enabling various navigational options, including querying

---

[4] One the earliest and prominent examples of knowledge panels on the web is the Google knowledge panel (Sullivan 2020).

and browsing, offering multiple cues to the user to help them assess the relevance of search results, and "fast navigation through connected space." In the model of user tasks defined in the IFLA Library Reference Model, the explore task is an open-ended exploration which can entail browsing and exploring relationships between entities. Discovery system support for this task involves "making relationships explicit, by providing contextual information and navigation functionality" (Riva, Le Bœuf, and Žumer 2017, 16).

The foil to known item search, serendipitous discovery, has been more difficult to facilitate in an online discovery system. As Intarapaiboon and Kesamoon (2019) point out, the user experience for unknown item searching in library catalogs is sometimes lacking and could benefit from browsing aided by domain ontologies. While rooted in a common refrain of library users seeking to recreate shelf browsing or coming across materials in a physical space, serendipitous discovery in the online context has often been ignored. Online experiences that feel serendipitous are often based on web analytics and tracking data and thus embed particular biases. One common feature in e-commerce interfaces is the suggestion of other items based on purchase history or aggregated social media data. Use of such external data sources would represent uncomfortable and potentially unethical directions for libraries.

At the same time, external data sources may include information that can help users better understand whether library resources are relevant to the ideas or people or subjects they are researching. These data sources may also capture contextual information that is not currently represented or may not belong within the library catalog itself but could help users better identify the relevance of the information they are viewing in the catalog. For example, a library catalog may choose to not include information about an author's professional or educational affiliations within the library catalog or authority metadata, but this information may be present in Wikidata and can provide useful context around an author. Similarly, subject heading authorities may not include descriptions, but Wikidata or DbPedia descriptions for these subject headings may provide useful context for end-users (Mak et al. 2021).

In external data, relationships between authors, subjects, and works may be captured at a different level of granularity than what is currently expressible in the MARC records. Linked data sources can encourage users to discover authors influenced by one another or by particular works, or authors who studied together or taught together. Linked data sources can reveal social relationships between people, places, and works. These are the types of connections that online search behavior reinforces when not limited by a MARC record context. Providing easier access to and navigation through these relationships and points of interest can help support open-ended and exploratory search.

Minding the Gap

The vocabularies and headings in use in the library catalog do not always map to users' own language or understanding of the concepts related to their search (Fresnido and Barsaga 2020; Trapido 2016; Wells 2020). As Allison (2010, 377) notes, "researchers may lack the vocabulary used in metadata schemes." Cuna and Angeli (2020, 506) state that "novice users are typically unaware of or unfamiliar with the MAchine-Readable Cataloging (MARC) structure, subject headings and classification numbers underpinning subject access points in library catalogs." Boettcher's (2020) user studies with undergraduate and graduate students also revealed that the vocabulary mismatch between user and catalog prevented discovery of relevant materials.

Although opportunities exist for training library catalog end-users to better employ existing classifications, vocabularies, and library catalog features such as facets to search for resources, we should consider any and all improvements in design that can mitigate the need for this kind of training. Golub (2018) defines multiple criteria for enabling better access to subject headings within a library catalog, including linking subject access points to resources, providing autosuggest, supporting browsing by individual concepts, and displaying broader and narrower information. Context to help disambiguate between authors can also be helpful (Boettcher 2020). The use of linked data presents opportunities for providing users with additional context around authors and subjects and to support users in more easily viewing and navigating to related and relevant resources and entities.

Our Approach

As part of the Linked Data for Production: Pathways to Implementation (LD4P2) and Linked Data for Production: Closing the Loop (LD4P3) grants, we undertook several prototyping experiments to explore the integration of linked data into library catalog discovery. We conducted user research preceding any design or development as well as usability evaluations for mock-ups or functional prototypes for each of these experiments. One of our driving questions was: how does the use of external data support particular open-ended discovery use cases? We also had to consider the following questions around data: which linked data sources exist that provide relevant information? Which of these sources can be connected back to the

metadata in the library catalog using identifiers or URIs? We recognize that additional questions around ethics and accessibility are also important and will discuss those later in this paper.

Each of the areas we explored showed some promise as well as opportunities for improved design or challenges around the integration of external data. For the work we describe below, we generated multiple prototypes to identify how to use linked data sources to support specific user tasks and then evaluated those prototypes with users. We focused on integrating linked data into existing library catalog technologies such as Blacklight without requiring library metadata be modeled semantically. This approach enabled us to experiment within the current library discovery paradigm while still providing insight relevant to implementing semantically modeled library metadata in the future. Below, we describe user research that preceded our explorations and then discuss our experiments in detail.

Preliminary User Research

Before we began any ideation around discovery solutions, we interviewed undergraduate students, graduate students, and faculty to understand how they use current discovery interfaces such as the catalog, specialized databases, or general internet search engines (Usong 2019). By listening to how they described their approaches to discovery and their understanding of information sources, we began to see the dichotomy between specialized and undergraduate researchers. Each had their own set of challenges and requirements.

For the specialized researchers, which included graduate students and faculty, discovery of new information sources such as dedicated databases and repositories was just as important as discovery of the actual materials. They value the discovery of new resources because those resources could lead them to other pathways as well as help them understand if others have already explored the area they are researching. Cross-references can also be helpful because they provide some much needed context; however, because this additional information is also labor-intensive to provide, it is not usually available.

Before undergraduate researchers can begin their research, they need to find background material to help them understand the subject matter better. The undergraduates we spoke with began their search in Google because they could do research on the internet first and then, armed with the information they found, be more successful finding materials in the library catalog. In their general internet search, they can find the correct vocabulary to use in their library catalog search. However, some do not know that specificity matters and thus experience challenges. For example, an undergraduate student assigned a paper on Egyptian art might need to know about "sequence dating" and "faience" to discover the material they need. They also need greater error tolerance in discovery systems because of their unfamiliarity with the subject matter and how the catalog works. This problem is compounded when material is in multiple languages using non-Latin characters.

One commonality between specialized and undergraduate researchers is the mismatch between the catalog vocabulary and the user's natural language. Even after years of research experience, this dissonance still presents barriers to discovery for specialized researchers. Undergraduate researchers have the additional problem of not being familiar with the subject matter and also having issues with misspellings. As a result of this research, we decided to explore mechanisms for providing context, relationships, and suggestions for undergraduate researchers. These mechanisms can still help specialized researchers as they explore domains with which they are not familiar as part of interdisciplinary research (Boettcher 2020).

Context Is Key

Traditional library cataloging has focused on controlled vocabularies,[5] which often do not provide enough information for users to confidently navigate our collections. All of our linked data discovery-related explorations have a common element—the introduction of context beyond traditional library information in a discovery layer.[6] In reflecting on each of these explorations, we seek to expose the different ways in which these discovery layers present information to users and center the UX research that guides our thinking beyond the modes of traditional library inquiry. Rather than simply asking whether linked open data makes something possible, we are allowing user feedback about whether something is useful, helpful, or meaningful to guide our explorations and center UX research as part of the design and development process. We also consider the impact of each exploration, including how each exploration influenced subsequent explorations and how they can help to reframe broader conversations about creating and maintaining metadata standards that center user experience and use of library metadata on different discovery journeys.

---

[5] Controlled vocabularies are curated, unique terms that are used to refer to people, places, concepts, etc. and are used in cataloging.
[6] This might be seen as an attempt to support the Functional Requirements for Subject Authority Data (FRSAD) "explore" user task (IFLA Working Group on the Functional Requirements for Subject Authority Records 2010).

The inclusion of knowledge panels was our first but not only exploration of incorporating contextual information from linked data sources into the library catalog. As part of this exploration, we asked the following questions: can we help users, when viewing search results, better assess whether the author or subject is relevant to their search? Can we help users see which other authors or subjects are related to the author or subject of the work they are reviewing and how they are related? Can we then provide related library resources that could support users in accessing additional materials which could help them in their search? Some of this context may be available in our library catalogs, but external data sources can provide more specialized or more expansive connections. These motivating questions helped frame the work below.

Knowledge Panels

Search engines like Google provide information boxes highlighting brief information about people and subjects alongside search results. These information boxes are commonly called "knowledge panels." Existing library catalog systems, such as the University of Wisconsin-Madison, incorporate external data sources such as Wikidata and DBpedia to provide additional information about authors of or contributors to library resources (Meyer 2016; Allison-Cassin 2019). Inspired by these existing systems, we generated multiple versions of knowledge panels in the LD4P2 grant and are working on a streamlined version in LD4P3.

Teams at Cornell and Stanford explored the design and development of knowledge panels which could provide descriptive information about authors and subjects (CUL 2020; Cramer et al. 2019; Khan, Worrall, and Skinner 2020b; SUL 2020). These panels are intended to support users in retrieving information about specific authors or subjects without leaving the search results or the item view page and in navigating to related authors, subjects, and works using the information displayed in the knowledge panel.

Figures 1, 2, and 3 show two different mechanisms of triggering a knowledge panel from prototypes developed at Stanford and Cornell during the course of the LD4P2 grant. In the first example from the Stanford LD4P2 prototype, the user has selected an author facet (Figure 1). On the results page, information from Wikidata about the selected author, including an image, description, notable works, and occupation, is displayed in a knowledge panel. The second example shows that the author knowledge panel for Beethoven includes musical clips from his work (Figure 2). In the example from the Cornell prototype, the user clicks on the info button next to the author for *Lincoln in the Bardo* and sees a knowledge panel incorporating similar information from Wikidata as well listing results from library digital collections (Figure 3).

In addition to author knowledge panels, we also experimented with knowledge panels for subjects and locations. Figure 4 shows a screenshot from the same page as Figure 3, which shows information about the novel *Lincoln in the Bardo*. This novel has the subject area "Lincoln, Abraham, 1809-1865." Clicking on the
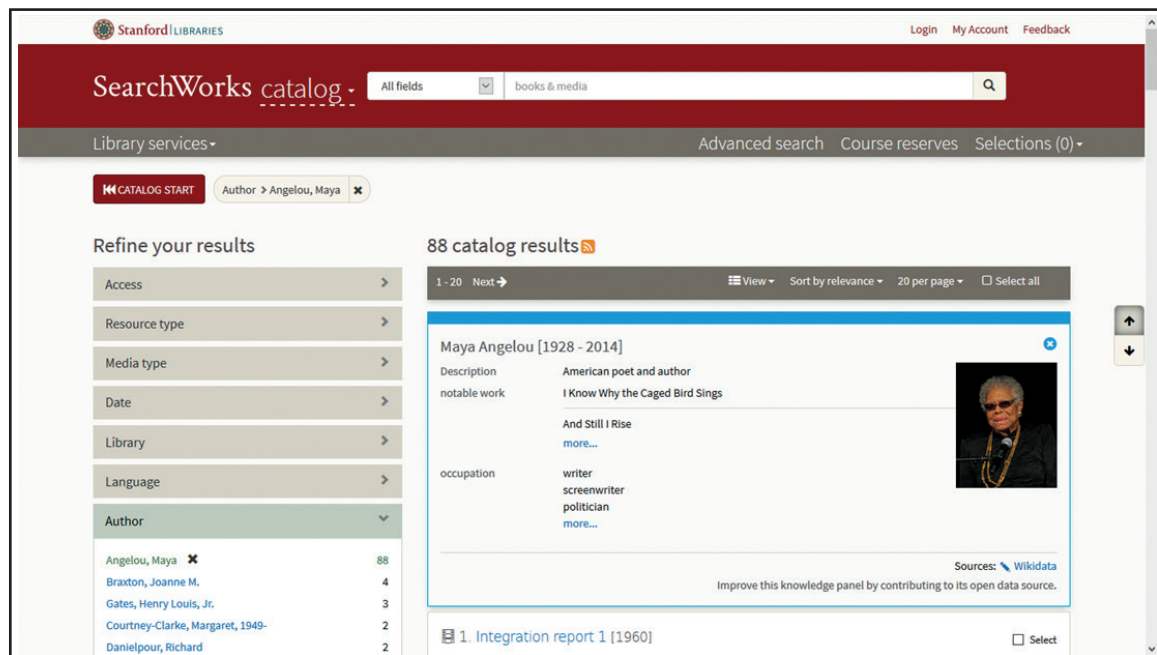

Figure 1: Stanford prototype showing author knowledge panel for Maya Angelou.
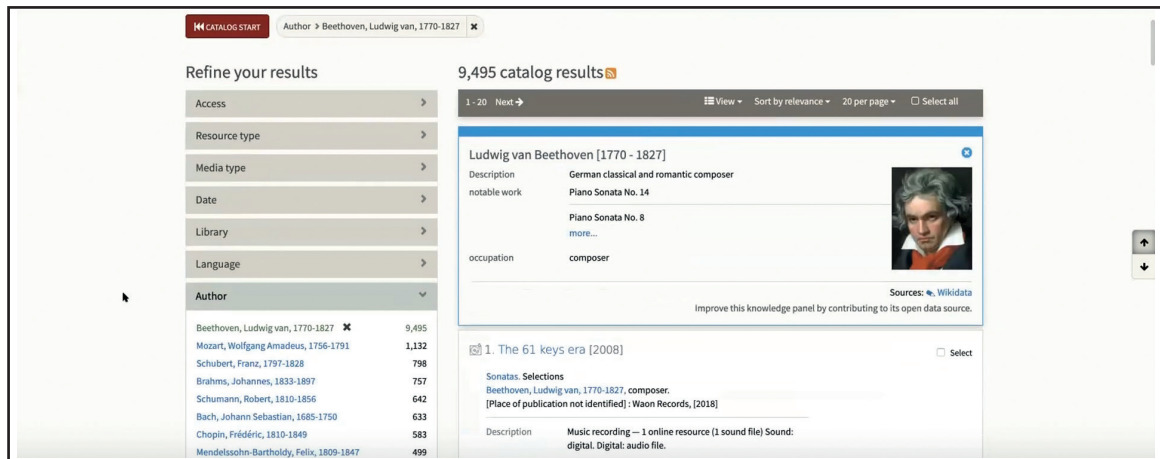
Figure 2: Stanford prototype showing knowledge panel for Ludwig van Beethoven, which includes musical clips.
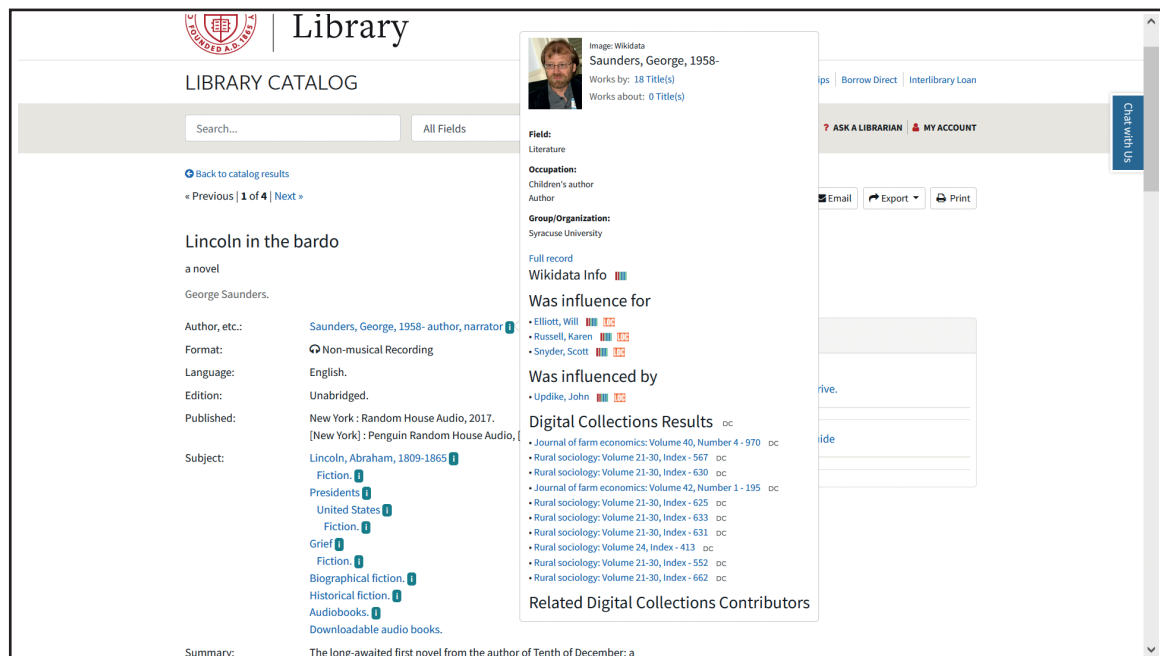

Figure 3: Cornell prototype showing author knowledge panel for George Saunders.

info button next to the subject opens up a knowledge panel which pulls in an image from Wikidata as well as related digital collection results.

Figures 5 and 6 show two different examples for incorporating knowledge panels centered on geographic information. In Figure 5, an example from the Stanford prototype, the user has selected the region facet value for "Palo Alto." The search results show the location knowledge panel for Palo Alto, bringing in descriptive information from Wikidata and coordinates from the gazetteer Who's On First to enable the display of a map view of that location. In Figure 6, an example from the Cornell prototype, information about the narrative location of the novel *Finnegans Wake* is retrieved from Wikidata. Geographic coordinates are retrieved from Wikidata to enable the display of a map view of the location in a geographic knowledge panel.

Data and Implementation

We utilized data sources that could provide information related to authors, subjects, and locations in order to provide context around library resources. All of the examples for this phase of exploration relied on client-side queries to retrieve information from the external linked data sources in real time and then display that information in the page. In other words, these queries were executed when the knowledge panel was

Figure 4: Cornell prototype showing subject knowledge panel for Abraham Lincoln.

Figure 5: Stanford prototype showing location knowledge panel for Palo Alto.

triggered within the web page. Stanford examples utilized URIs added to their MARC records for these experiments. The Cornell examples executed queries that first retrieved the LOC entity local name (i.e., the last part of the URI) by searching against the LOC Linked Data Service, id.loc.gov, and then conducted
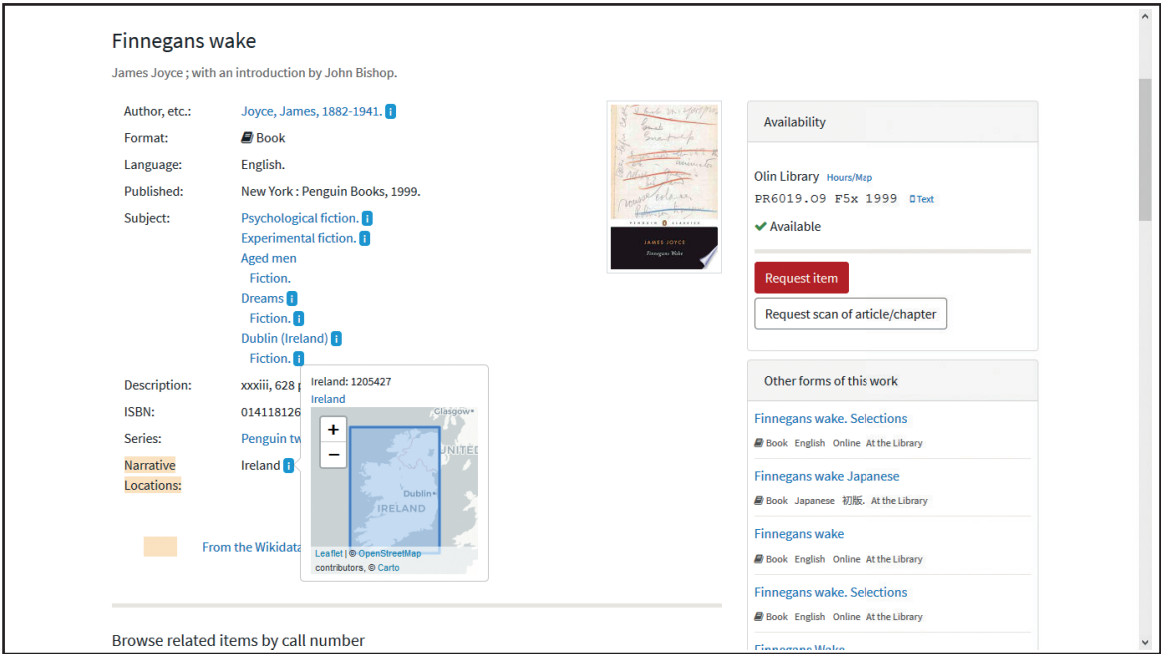
Figure 6: Cornell prototype showing narrative location popup for Finnegans Wake.

Table 1: Data sources used for knowledge panels

Source	Used for
Id.loc.gov	URI based on string for Library of Congress Subject Headings (LCSH) and Library of Congress Name Authority File (LCNAF)
Wikidata	• Author and subject image • Notable works • Influence relationships • Awards received • Narrative location • Geographical coordinates for location
Cornell digital collections	Search results based on keyword for author or subject headings
Who's On First	Geographical coordinates for location

a second search against Wikidata to find relevant connections. The first search against id.loc.gov utilized a text string search which also took care to remove trailing spaces, hyphens, and periods to ensure better results. The resulting LOC URI was then used to match against possible URIs in Wikidata. Having any URI connections already saved within the Solr search index would allow for more easily obtaining identifiers to retrieve entity level information dynamically within the page. Table 1 combines the data sources used by the Stanford and Cornell prototypes and lists which information was retrieved from these sources.

User Research

Mock-ups and prototypes for the work shown above were evaluated at both Stanford and Cornell in separate studies. High-level results from these evaluations indicated that users perceived the following as useful: knowledge panels that represent contextual information and related relationships that can lead to further searches in the catalog. Where users differed was on *which* pieces of information are relevant for context or further searches. For example, there were differing perspectives on the usefulness of including links to books by authors who had won one of the same awards as the author whose knowledge panel was being viewed. The integration of digital collections information was regarded positively, although not all mock-up evaluation participants at Cornell knew what content was entailed by digital collections. Future work could focus on differences between knowledge panel layouts (e.g., within page, visible by default, or available on clicking) as well as continued exploration of which properties and linkages within the knowledge panels

would be considered useful. There may also be a connection between knowledge panels and browsing, where browsing could occur across a collection of knowledge panels.

The Stanford team evaluated both mock-ups and interactive prototypes using a "guerrilla user testing" approach where they held short focused sessions with students at the library. They recruited ten participants: eight undergraduate students and two graduate students. They found that all participants thought the author panel was helpful. One participant stated that they did not need the panel but was not opposed to its inclusion. For the remaining participants, the knowledge panel provided the author with credibility and could potentially lead them to other relevant information. (As one participant said, "it's the bombdot-com.") Most participants did not search for media so were not as interested in the media knowledge panel. One participant who looked for music in the catalog thought having the track list would be helpful. The rest said that they usually go to other websites to look for media.

Prior to finalizing the Cornell knowledge panel prototypes, Cornell and Stanford collaborated on the creation of mock-ups which were used in evaluation with Cornell students. Three Cornell undergraduate students as well as a graduate student and a postdoc participated. They provided feedback after viewing the mock-ups and having the associated workflows and tasks described to them. In this set of evaluations, knowledge panels were generally found to be useful. Transitions between search results or terms from external sources and catalog results appeared to make sense to participants. The integration of the digital collections source was considered useful.

Participants showed varied opinions regarding which properties included from Wikidata were useful. Three participants appreciated the inclusion of "awards received" but two noted that they were not interested in this information. None of the participants objected to the inclusion of influence relationships and three noted that they liked the inclusion of this information. Two participants asked questions about how this relationship was assessed or determined, with one asking for an explanation and another for sources such as interviews.

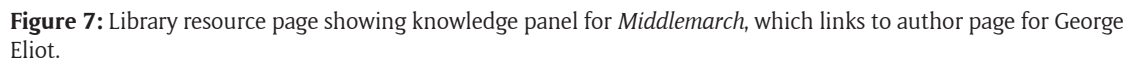### Impact on Library Systems and Data Practices

This exploratory work on the integration of knowledge panels relies on a commonly utilized framework for sharing information online and brings external information through that framework into the context of the library catalog. By using a familiar pattern, this approach meets user expectations for interacting with external data and separates that information from the library-controlled catalog information in a visually distinct way. Adopting this approach allows user research to focus on which data points are most valuable and usable, reinforcing the goal of bringing in external data to contextualize and advance existing inquiry. Given the lack of consensus on which properties are most useful and the need to consider streamlining knowledge panels, future studies might consider evaluating patron responses to different properties for different entity types across available datasets. With this information in hand, we can better determine which properties to maintain and which datasets to make explicit connections to.

Author and Subject Pages

Various library catalog implementations include separate browse functionalities for authors and subjects. These browse indices often reflect or relate to the library classification and authority information used within the pages displaying resource information as well as within facets used to search by author or subject. The primary motivation for introducing additional contextual information about authors and subjects is to help users assess and disambiguate which author or subject they wish to explore. As a result, we have prototyped and assessed views on authors and subjects that bring together information from across the library catalog and related library sources as well as from external linked data sources. We have experimented with blending these indices with item views, thus bringing together known item search support with more open-ended browse functionality.

Where knowledge panels were intended to provide a quick overview of information about authors and subjects and add additional external data points to library catalog information, author and subject pages are intended to support users in obtaining a more complete view of library and external data associated with a particular author or subject. In addition, the information on the page is intended to help show how a particular author or subject relates to other authors and subjects, thus situating a single entity within a larger network of relationships.

During the LD4P3 grant, we created prototypes for bringing together information about authors and subjects into dedicated pages. The first example, in Figure 7, shows what the user would see looking at the Cornell library page for the novel *Middlemarch* by George Eliot. Clicking on the "info" button displays a simplified version of the knowledge panels we discussed in the preceding section. Clicking on the "full record"

Figure 7: Library resource page showing knowledge panel for *Middlemarch*, which links to author page for George Eliot.

Figure 8: Author page for George Eliot showing information, library resources, and publication timeline.

link takes the user to the author page for George Eliot. This page uses Wikidata information to display descriptive information about George Eliot, a timeline displaying works by George Eliot, and a graph showing people who influenced and were influenced by the author (Figures 8 and 9). Links to library holdings and related resources are available in the right-hand section of the page.

We also created subject pages. In the example below, the user can click on the "info" button next to a subject listed on the page for a book (Figure 10). Clicking on "view full record" will take the user to a page dedicated to the subject "Russo-Japanese War, 1904-1905." This page brings in descriptive information
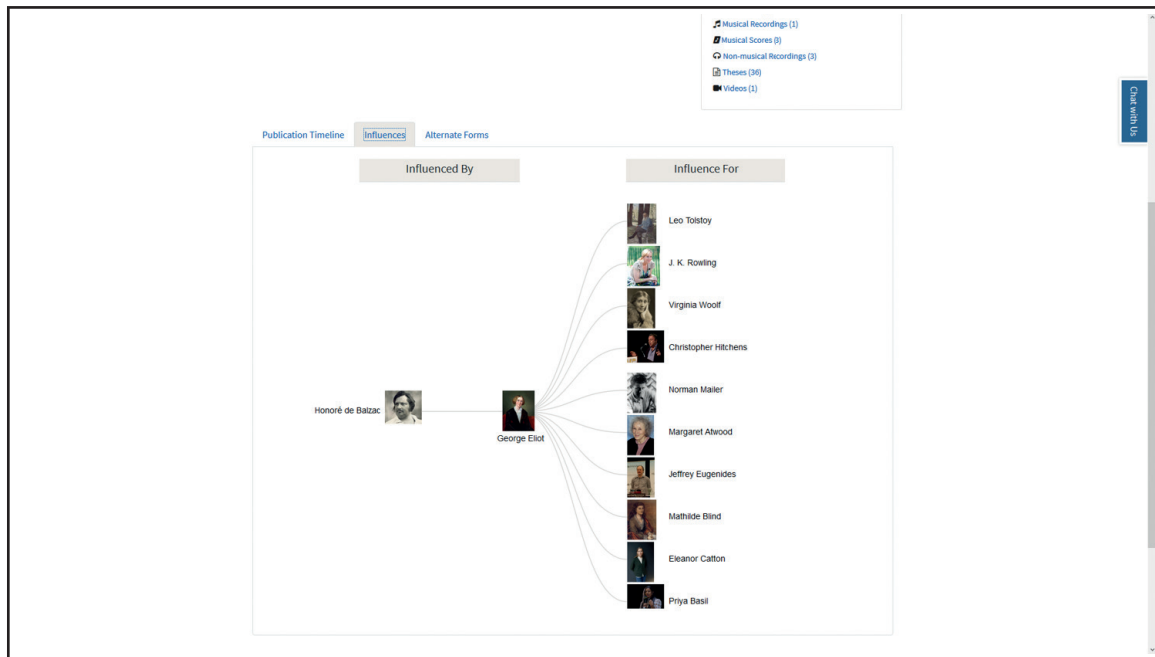
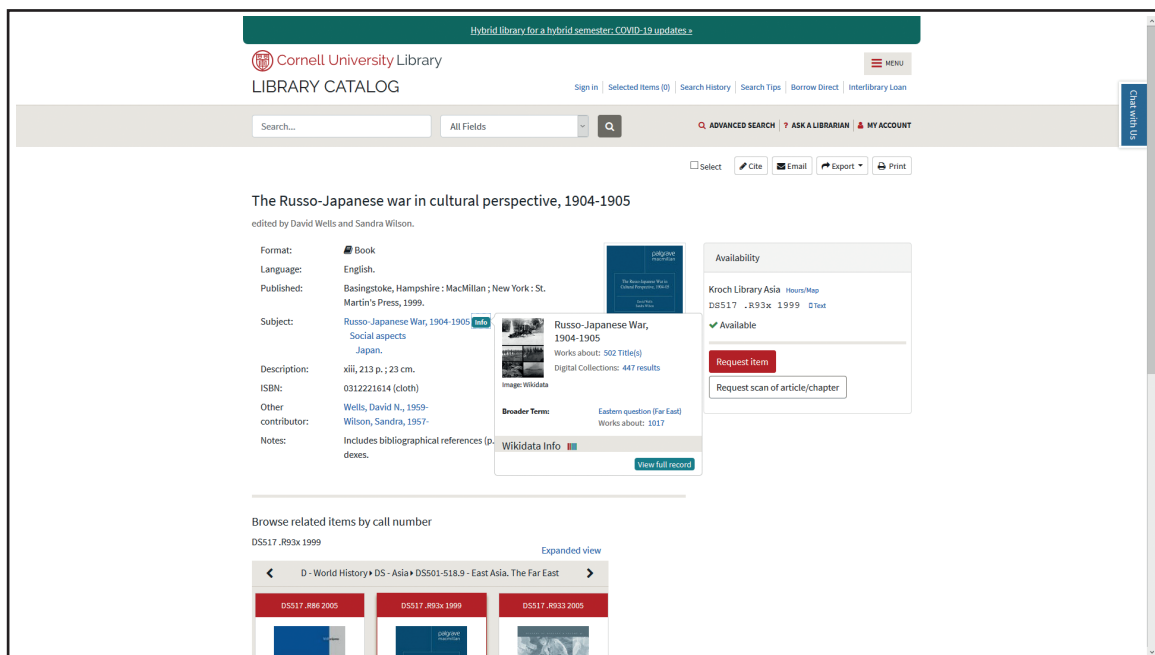**Figure 9:** Author page for George Eliot showing influence relationships.

Figure 10: Subject knowledge panel for *The Russo-Japanese War in Cultural Perspective, 1904–1905*, which links to subject page for "Russo-Japanese War, 1904-1905."

about the subject, related temporal and geographic information, broader and narrower subjects, and links to related call numbers using a variety of sources including Wikidata, PeriodO, Library of Congress Subject Headings (LCSH), and Library of Congress Classification Numbers (LCCN). The first few related library catalog results for this subject are included. Temporal and geographic information about the subject being viewed as well as about broader and narrower subjects is used to populate the timeline and map view. Clicking on the "display all subjects" checkbox populates the timeline and map so the user can see when other events may have occurred or the temporal coverage of chronologically similar subjects (Figures 11 and 12).
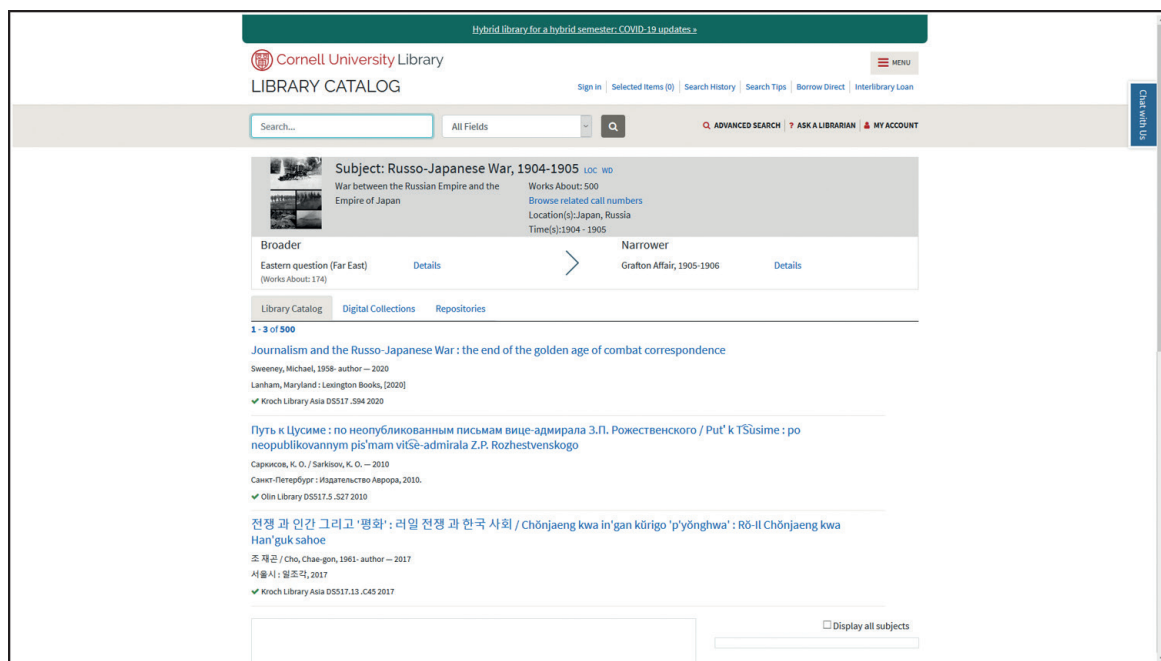
Figure 11: Subject page for "Russo-Japanese War, 1904-1905" showing broader and narrower relationships and related library catalog resources.
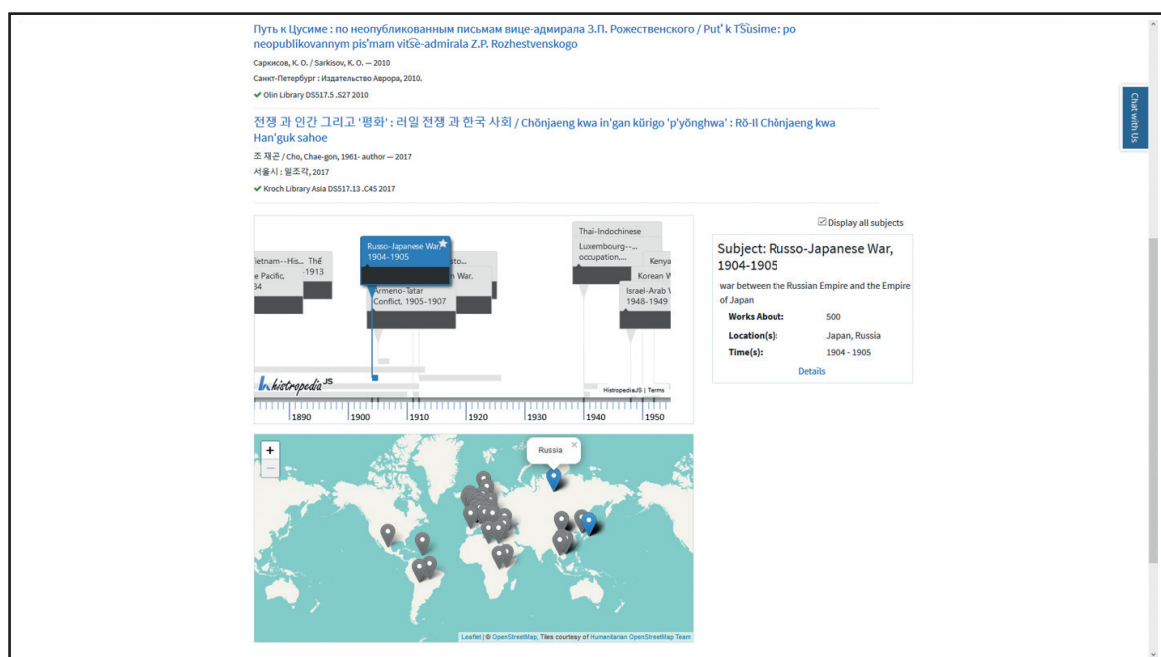

Figure 12: Timeline and map on subject page for "Russo-Japanese War, 1904-1905."

Data and Implementation

The author page relied on multiple client-side real-time queries that utilized information from Wikidata as well as from library catalog bibliographic and authority metadata. The information driving the subject page was retrieved from multiple data sources and combined into separate Solr indices, enabling quick queries to drive the timeline and map and to show links to related call numbers. The page also used client-side queries to retrieve geographical coordinates for Wikidata URIs representing locations as well as to display broader and narrower subjects for a particular LCSH URI. Table 2 below summarizes the data sources and their use in the prototypes.

Table 2: Data sources used for author and subject pages

Source	Used for
Id.loc.gov queries or data downloads for querying and indexing	• URI based on string for LCSH and LCNAF • LCCN related to LCSH URIs • Broader and narrower relationships for LCSH • Temporal and geographic components for LCSH URIs
Wikidata	• Author and subject image • Works for publication timeline • Influence relationships and related information • Geographical coordinates for location
PeriodO	• Temporal and geographic information related to LCSH URIs • Wikidata URIs for locations related to LCSH URIs

User Research

We recruited five undergraduate students to evaluate this prototype. With each student, we scheduled a half-hour Zoom session to conduct a think-aloud session in which they were asked to complete a set of tasks for finding information on the author and subject pages described above.

The results showed that the page design enabled participants to find most of the information they were asked to retrieve. When asked about which features from the prototypes they would find useful, all participants noted at least one feature related to subjects would be useful. Three said the author timeline was useful and two thought that alternative forms of names would be useful. When asked about what kinds of search tasks users perform in their own work, all mentioned some use of subjects, including related subjects and broader and narrower subjects. Three participants mentioned that they look for or use information related to authors, such as author suggestions or author birth names.

Although we were not planning on testing or evaluating the design or placement of the "info" button or "view full record" link, our user testing showed a need to clarify the role of the info button with respect to the author link next to it. Labels for certain terms, such as "alternate forms" as seen on one of the tabs in Figures 8 and 9, could be made clearer. Although participants were able to use the timeline to find subjects, some participants noted confusion around what was being displayed on the timeline. Future improvements could thus focus on clarifying the role of the timeline and map on this page.

Impact on Library Systems and Data Practices

In this example, incorporating external data presented a fuller context for specific authors or subjects. Rather than relying on a demonstrated pattern for connected information, this approach demonstrates the breadth of what additional information is available for contextualizing discrete persons or concepts online. Bringing that content to the forefront enables the catalog to go beyond a common linear behavior of A–Z heading browse to a multidimensional approach in normalizing the aggregation of data sources outside a library catalog and crosswalking between them. As previously mentioned, how best to provide intuitive user controls to invoke the contextual author and subject pages or to present library resource result sets related to those authors and subjects remains an open question; with more consideration, a successful design pattern might emerge beyond the combination of info button and click-to-search heading. The author and subject pages, with their expanded scope, allow for greater browsing than streamlined knowledge panels; still (similar to knowledge panel findings), we can and should continue to determine which datasets and properties provide the greatest value for browsing and make concerted efforts to connect to and/or capture this information.

Music, Artists, and Albums

The crowd-sourced Discogs platform is a rich source of high-quality structured data about music, artists, and albums that music catalogers frequently use as a reference. As a marketplace, the Discogs community has a vested interest in accurate data in order to assure successful sales transactions of recordings when a collector is seeking to make a purchase from a vendor. The precision of these descriptions is on par with how libraries hope to describe their collections, though we often do not have the cataloging resources to capture such rich data. On occasion a cataloger will copy and paste information from a Discogs description into a MARC record.

Our work sought to discover whether we could make trusted dynamic connections between Discogs and MARC descriptions to save cataloger time and have a more far-reaching impact across our recorded music collections.

After evaluating the Discogs API, the Cornell team developed an integration that uses existing identifiers and other metadata in library records to retrieve additional information from Discogs. This information is included in the catalog item display, and in many cases it significantly enriches the item description. Unlike with the knowledge panels, we chose for the default display to be a seamless integration of the Discogs data as if it were native to the library record; to make clear which data are external, a link is provided in the interface to highlight the Discogs data (Figure 13). To adhere to Discogs' API Terms of Use (2018), we also provide a link out to the corresponding Discogs release page.

Data and Implementation

The Discogs API is the main source of information used to query and retrieve information, as described below. While not RDF-based linked data using URIs to link concepts and entities, the Discogs data nonetheless richly models information about recording artists and releases and offers robust APIs for access. Library

Figure 13: Discogs information highlighted on library page for the musical recording *You're the Guy I Want to Share My Money With.*

catalog pages incorporate Discogs data using a strictly client-side approach, where real-time queries are made against the Discogs API when the page loads. A call is made to the Discogs API using either the Discogs identifier for a release (if present in the catalog data) or a query that is generated using information from multiple fields in the metadata, including title, record label or publisher, and artist. Contributors, notes, track lists, and publication information from Discogs are then added to the item view when a match is made.

User Research

During a lively focus group of library staff, including stakeholders who maintain Cornell's Blacklight catalog, we received decidedly positive feedback on the Discogs feature. After we walked the large group through the functionality and underlying mechanics, staff were pleased that we were using such a "highly curated" dataset. Participants found track lists to be very useful in understanding whether a record was of interest, and we hope to index track titles as well as other Discogs data in the future to allow for better searching. Further, participants asserted that the data being brought in "doesn't feel like extra data," but rather enhances the existing description. There was enough agreement on the usefulness of Discogs integration to bring this feature to production.[7] In the future, we can also get targeted feedback on this feature from library users who interact with music materials to supplement staff feedback.

Impact on Library Systems and Data Practices

While limited to the music context, this exploration highlights the impact of subject-specific behaviors and preferences on where and how external data are deemed helpful or useful. Connections between pieces of music, writers, performers, and collections of music in a variety of forms provide a salient use case for the value in connecting across source catalogs. Going forward, we must consider whether our approach is the best way to cite external data when other sources (or more than one source) of external data are used to enhance a resource description. Stakeholders voiced valid concerns about inaccurate matches, leading to the development of a procedure where a cataloger can override a match by adding a Discogs identifier directly to the MARC record.

Browsing

McKay, Buchanan, and Chang (2018, 348) point out how browsing "suffers from a definitional problem" and refer to the following as a good working definition: "Bates (Bates, 2007) notes that browsing is the viewing of a large and interesting scene, and the identification and sequential (rather than concurrent) examination of objects of interest within it." They discuss how users browse library shelves with serendipity and physical co-location affecting what they view or borrow.

For our work, we focused on providing users with the following browsing support:

- Enabling browsing by displaying top-level categories or collections that then provide navigation opportunities to what is contained within those categories or collections.
- Providing users with the option to switch between or interleave browsing and searching tasks.

During LD4P2, we experimented with various browsing approaches, including the design and development of mock-ups and prototypes for author and subject browsing (Khan, Worrall, and Skinner 2020a). The timeline below (Figure 14) uses Wikidata and Library of Congress Name Authority (LCNAF) information to display when an author lived and worked. Clicking on any of the timeline cards will populate a small knowledge panel on the bottom left, showing birth and death information and a short description. In addition, the user can see library catalog search results in the bottom right.

Figure 15 below shows an example from a Cornell prototype that uses LOC linked data to provide a subject-browse experience. The user sees this page after they have selected the top-level LOC Classification "Philosophy, Psychology, and Religion." The middle column shows the LCSH that relate to this classification. The page shows that the user has selected "Abbeys" from this list, which displays broader and narrower information as well as related library search results in the right-hand column. The left-hand column allows the user to further review related subject headings and results for subclassifications of "Philosophy, Psychology, and Religion." Figure 16 shows the catalog search results page, which includes a "browse" button next to the subject query "Abbey." Clicking on this button will take the user back to the browse screen, thus enabling the user to navigate to switch between browse and search views.

---

[7] Documentation on the Discog integration in Cornell Library's Blacklight Catalog is available at https://wiki.lyrasis.org/display/LD4P3/Production%3A+Discogs (archived at: https://perma.cc/H8GJ-53NL).
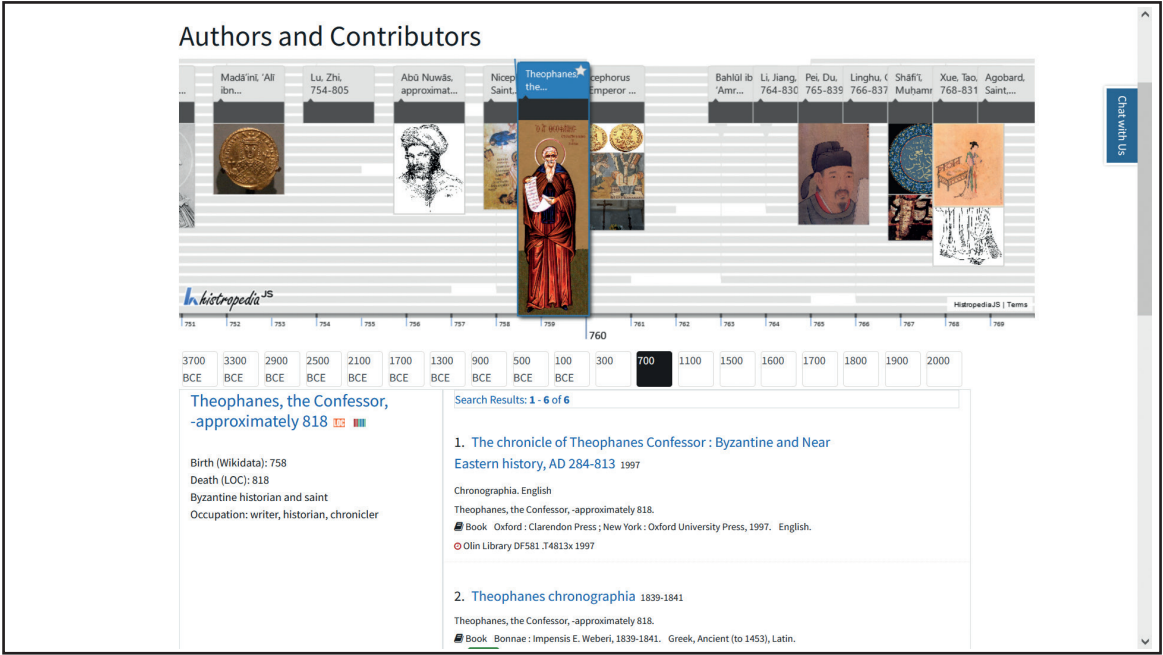
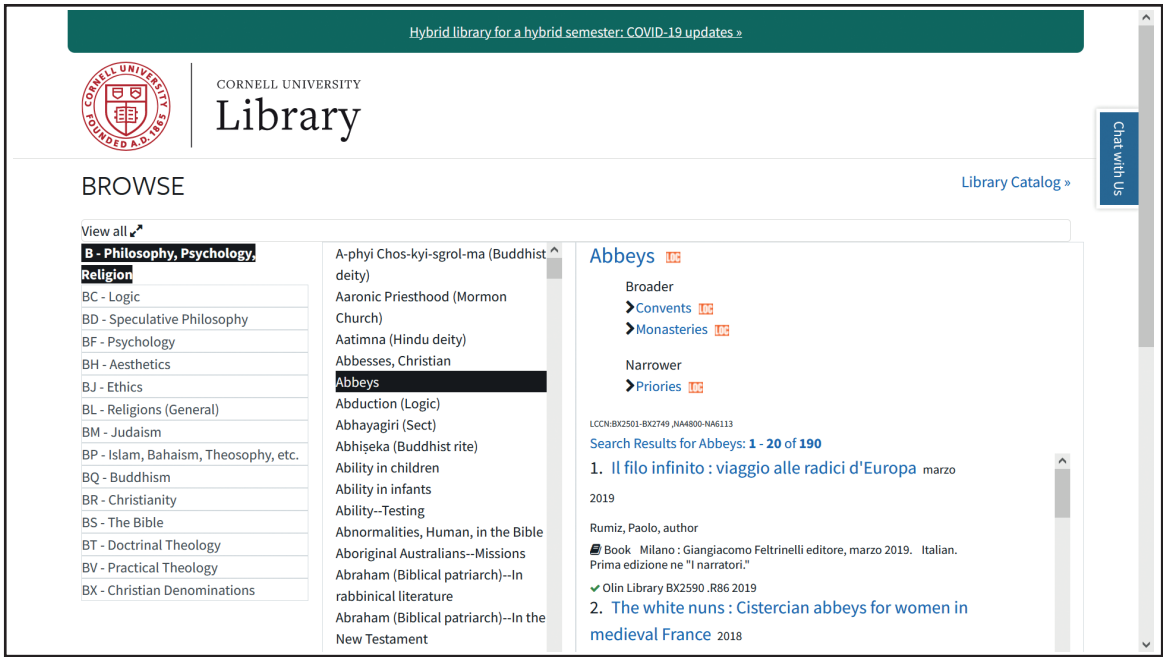**Figure 14:** Author timeline for Theophanes, the Confessor.

Figure 15: Cornell University Library subject browse page for "Abbeys."

Data and Implementation

For both author and subject headings, we created Solr indices to integrate information across multiple linked data sources and provide quick retrieval and querying for the front-end pages. The index driving the author timeline integrated information from Wikidata and the LOC for birth and death dates as well as start and end dates for activity. The 1,558,367 entries in this search index were retrieved from querying a triple store containing LCNAF information. Using Wikidata queries, 949,446 of these entries were updated with corresponding Wikidata URIs and with image URLs. The index driving the subject browse consists of 89,068 entries with LCSH labels, URI, and LOC Classification numbers and codes. Additionally, this index enables looking up which subject headings correspond to classifications that start with a given letter or given two letters. Table 3 summarizes how data sources were used for the author timeline and subject browse pages.
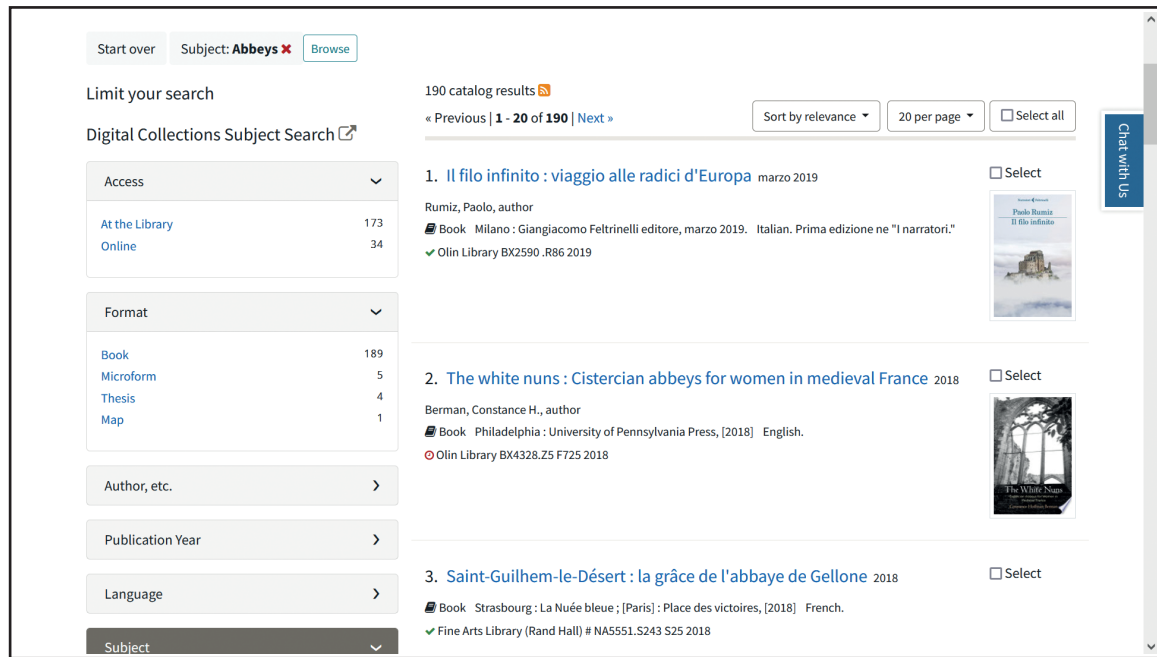
Figure 16: Search results page in Cornell University Library catalog showing browse button next to "Abbey" search query.

Table 3: Data sources used for author timeline and subject browse pages

Source	Used for
Id.loc.gov	LCSH and related LCCN
	LCNAF, including birth and death dates
Wikidata	Images for authors
	Description
	Occupation
	Start and end dates for authors' activity

User Research

Using mock-ups representing various browse options, two researchers recruited and interviewed six students in a Stanford University library lobby. Four of the students were first-year undergraduate students and two were graduate students. The researchers showed students mock-ups with different kinds of browsing experiences (Figures 17 and 18) and asked them questions regarding whether they would find the features helpful. Of these features, subject, call number, and timeline browse generated positive reactions.

Most of the students shared the sentiment that was expressed by a participant: "The more ways to search, the better." The graduate researcher was skeptical that we could provide a useful browse interface for her very specific needs. In the author browse section, participants found the knowledge panel to be helpful, but most would prefer to use subject browse. The timeline universally appealed to all types of researchers and could cover multiple use cases such as displaying the life of an author, date of publication, or the temporal setting of the narrative in a book.

Impact on Library Systems and Data Practices

When COVID-19-related closures prevented access to libraries and accessing materials was moved entirely online, a frequent concern raised by faculty was the loss of serendipitous discovery; they wished for us to reimagine the online experience to mimic physically browsing the stacks. While an exact replication of the physical browsing experience is not feasible, browsing online has distinct advantages to physical browsing, namely the inclusion of materials housed in different buildings as well as the mixing of electronic and physical materials. Explorations to uncover a better browsing experience are paramount to ensuring that library discovery environments are relevant as tools to facilitate research. Incorporating contextual data and new user interfaces as avenues to navigate between resources is a big step toward this goal.
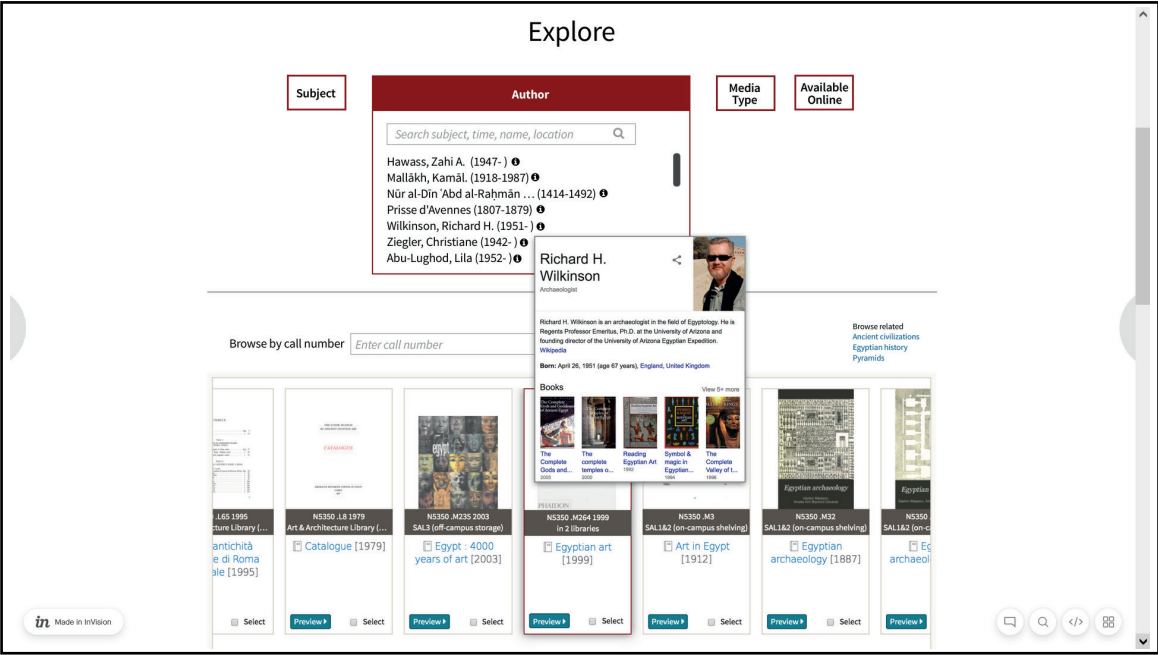
Figure 17: Mock-up showing author knowledge panel for Richard H. Wilkinson.
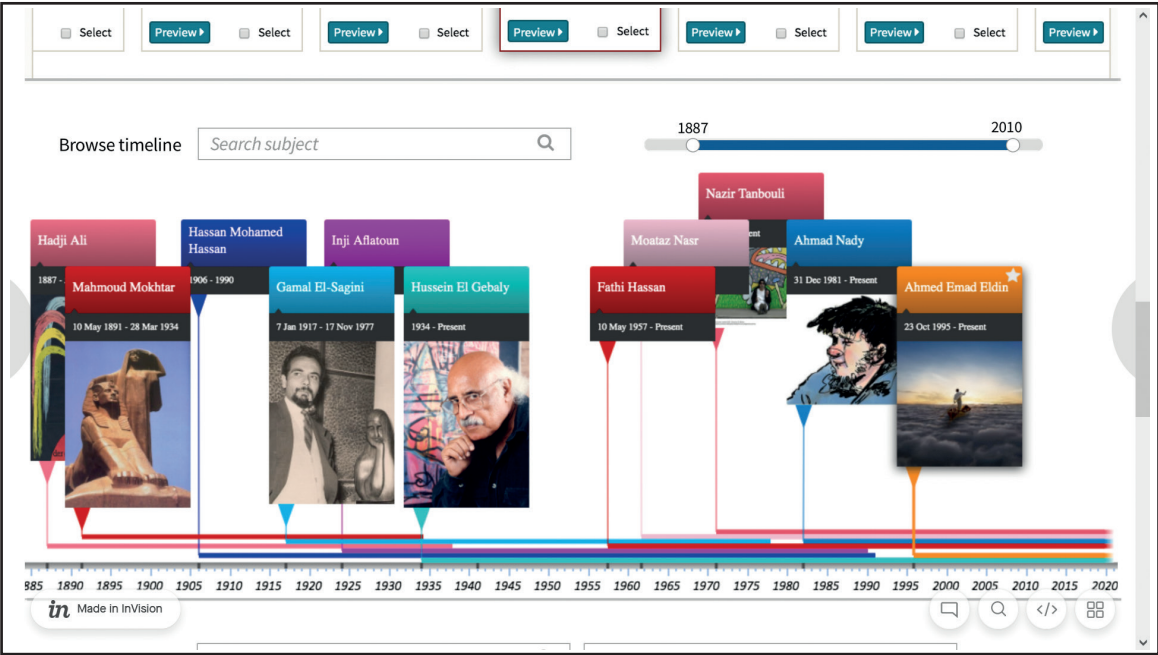
Figure 18: Mock-ups showing author timelines.

Suggestions

Trapido (2016, 18) noted that many topical search failures in library discovery stemmed from a "mismatch between the user's query and the vocabulary in the system's index." Similarly, author search errors resulted from various factors, such as a mismatch between the user query and the form of the name in the bibliographic record and a lack of system support for searching by variant versions of the authorized name.

Search engines often employ autocomplete to provide options for users once they have typed in a partial search query. Some library catalog systems have also taken on the challenge of providing entity-based suggestions to their users. For example, the University of Ghent Library search box provides suggestions for author and subject queries based on partial and complete user search query strings.

As part of LD4P2's exploration of the use of linked data to provide useful suggestions for end-users, we explored developing an autosuggest prototype (Khan, Worrall, and Skinner 2020c, 2020d). We wanted, when users begin typing into a search box, to display and offer them the chance to select names and labels that exist in our indices that will lead to relevant search results about specific entities such as authors, subjects, genres, and locations. If they type in a part of a subject or author's variant label, the system should be able to provide relevant matches in the catalog. Furthermore, if multiple author headings exist within the catalog that are linked together using "see also" relationships, the user's query for one should provide suggestions for these related headings.

Two examples below help illustrate the direction we took in implementing autosuggest. In the first screenshot (Figure 19), the user has typed in "twain." The autosuggest feature matches this string against the author "Twain, Mark, 1835-1910." A Wikidata description next to the author name provides additional context to help identify this result. In addition, the autosuggest feature shows that Mark Twain has a linked author heading "Clemens, Samuel Langhorne, 1835-1910," which will execute a separate query against the catalog. The user query also matches part of various other names for locations and subjects. The autosuggest shows suggestions where any word in the suggestion begins with the query, so both "Twain, David, 1929-" and "Missouri > Mark Twain Lake" are displayed as matches. Each suggestion also shows the library catalog search count. The second example (Figure 20) shows that the user has entered the term "heart attack" and autosuggest recommends the subject heading "myocardial infarction" because the query matches part or all of the variant label for "myocardial infarction." In this case, the variant label does not map to a separate subject heading. The matching variant label is displayed to confirm the reason for the appearance of this suggestion.

Data and Implementation

We first retrieved data from the library catalog author browse index and the subject, location, and genre FAST facet values from the main search index. We used id.loc.gov and OCLC FAST lookup APIs to retrieve URIs corresponding to these string headings. We then set up a separate Solr index using this data. We queried FAST and Library of Congress information to add variant labels and "see also" URIs. For authors, we queried Wikidata to retrieve description, image, and pseudonym text for particular Library of Congress URIs. This information was also added to the Solr index. We configured this index to enable matching the beginning part of words. Table 4 below summarizes the use of linked data sources for this feature.

User Research

We had two rounds of usability testing related to autosuggest. In the first session, four undergraduate students and one graduate student were recruited at a Cornell library. This session involved a review
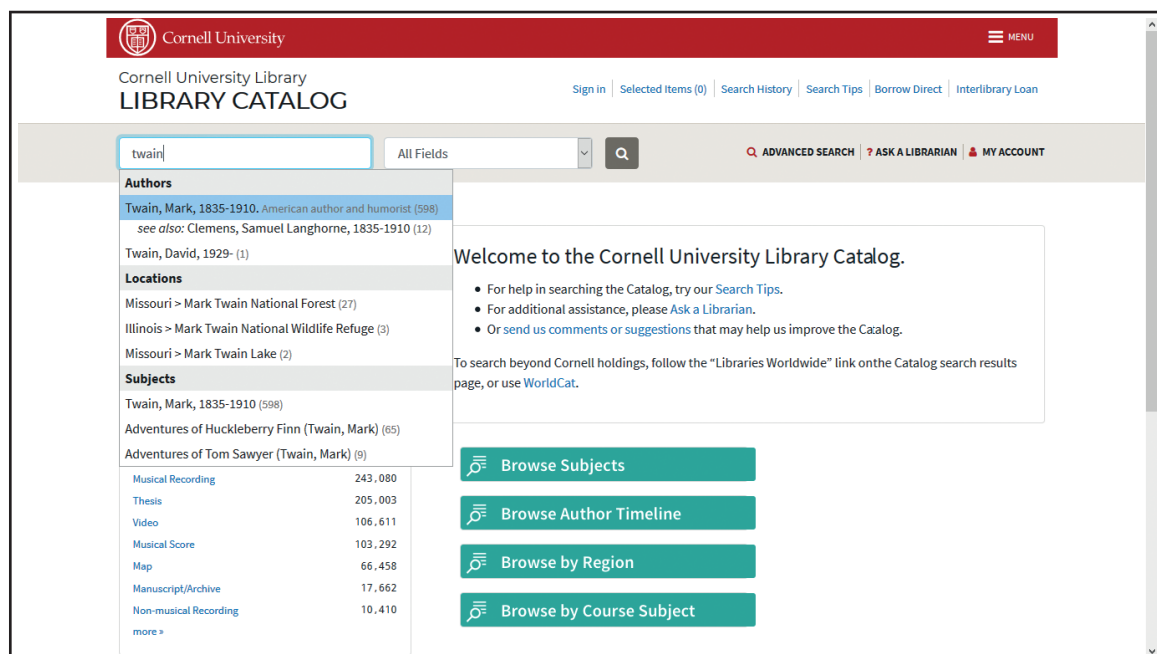

Figure 19: Suggestions for the query "twain" in the Cornell University Library catalog.
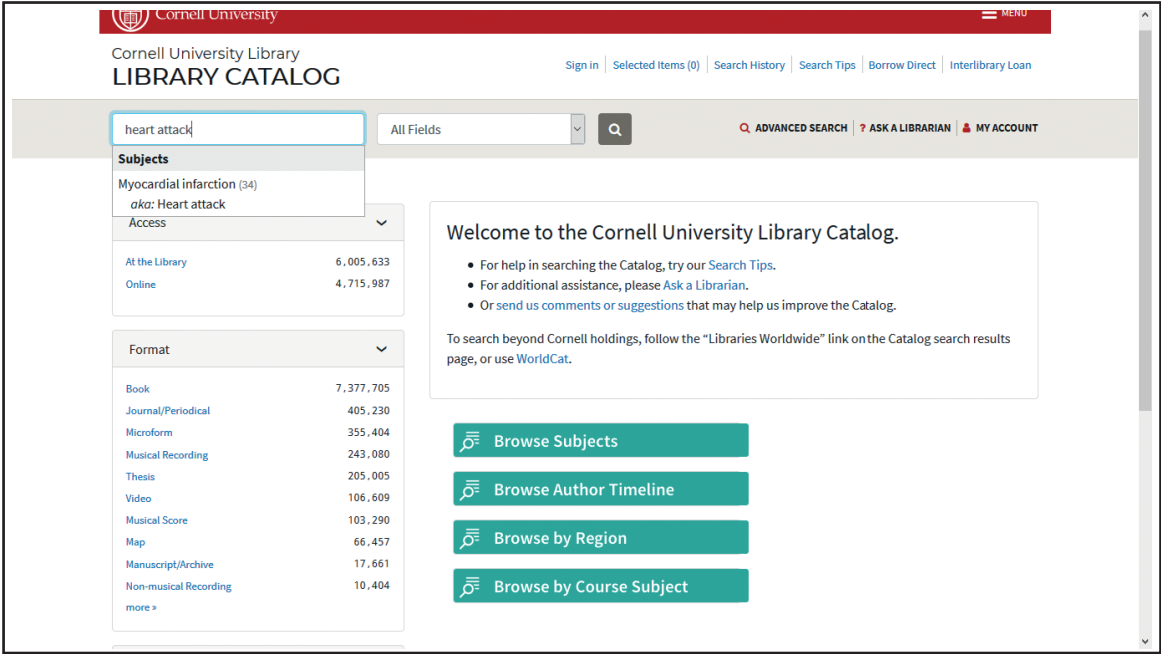
Figure 20: Suggestions for the query "heart attack" in the Cornell University Library catalog.

Table 4: Data sources used for autosuggest feature

Source	Used for

Id.loc.gov	• LCNAF
	  ◦ Variant Labels
	  ◦ "See also" URIs

FAST	• Variant labels
Wikidata	• Image • Description • Pseudonym text

of multiple features, including early autosuggest prototypes. Results indicated that participants thought autosuggest was a useful feature which could benefit from additional support for misspellings. Furthermore, the label "author" could be revisited to clarify what the list indicates (i.e., works by versus works about).

With the aid of the Cornell University Library Usability Working Group's expert facilitation and help, we conducted think-alouds with five Cornell University Library user representatives. In general, participants were able to use the information displayed in the autosuggest to distinguish between types of entities and to identify headings related to variant labels or connected using "see also" properties. Participants were able to use the descriptive text retrieved from Wikidata to distinguish between authors with similar names but different occupations. One participant did want to see if typing in the occupation along with the name would help in retrieving results, but we noted that we are not matching on occupation text.

For both authors and subjects, all participants were able to search for a variant label and find the preferred heading. Most participants understood what the term "aka" stood for in the results and thought this display of information was useful. When searching for an author, one participant indicated that they were not sure which heading was authorized since both the variant and preferred versions ended with date strings. For the shared pseudonym example, participants were able to find the names linked with "see also," but it was not clear to all that the searched name was a pseudonym shared between the linked names. Suggestions for clarifying this connection included using the term "pseudonym." Four out of five participants appreciated the resulting knowledge panel on the results page as a way of confirming the search they had conducted.

Impact on Library Systems and Data Practices

Given that many users are unsure which terms to use when beginning research, autosuggest has the potential to bring researchers to useful search terms pre-emptively rather than through a lengthy process of iterative searches and to obviate the need to train users in controlled vocabularies in advance of using our discovery layers. Users may also be able to evaluate different terms' efficacy and select useful terms more efficiently in light of the types, contextual information, or number of results displayed for different suggested labels. Users finding this feature useful suggests that catalogers should continue to link to and maintain authority records and entity descriptions to make this type of searching possible.

Questions to Consider in Production Environments

In the course of our experiments and evaluations, we have compiled recommendations reflecting different dimensions to consider when enhancing discovery through the use of external or linked data sources. We have summarized these recommendations in Table 5.

    While we have focused on using a research and prototyping process to explore our design and research questions, we are also in the process of defining and evaluating how to take some of these results and approaches into live production systems. Below, we describe some of the questions to consider when approaching integration of such work into production library systems.

Standards and Testing

What requirements exist for the following areas: accessibility, user evaluation, software, user acceptance tests? Have these requirements been met?

Data Dependencies and Acknowledgements

Which data sources provide information for these features? Are there requirements or data provider policies for when and how these sources of data should be displayed to and identified for users?

Data Unavailability

How will the system react if data are not available? For example, the page may be required to not display the sections dependent on external data if those data are not available, while still displaying the rest of the page.

Ethical Concerns

The use and display of data presents a host of possible ethical concerns. The LD4 Ethics in Linked Data Affinity Group (2021) has been exploring some of these concerns, which range from data being inaccurate or misrepresentative to potentially posing danger to the individuals being represented. While we will not

Table 5: Recommendations for decision-making and analysis in the areas of user research, data analysis, and implementation

User research	Data analysis	Implementation approaches
– Identify prior related research. – Identify or designate primary audience. – Define use cases or tasks motivating the design. – Describe the origin of use cases (e.g., stakeholder requirements, user research, etc.). – Find related interfaces (e.g., systems supporting discovery or displaying linked data information).	– Decide on relevant data sources. – Identify points of connection between the library catalog and data for discovery (e.g., identifiers or URIs in the MARC or from library authorities that can link to other data). – Identify the APIs or query services required to get data from linked data sources. – Identify whether the data source is open or proprietary.	– Decide or identify if information will be added to display alone and not to back-end systems. – Decide or identify whether information will be integrated with existing indices or back-end systems. – Identify whether data will be retrieved using real-time queries or API requests. – Define if there are requirements to implement new indices or back-end data storage systems. – Define if there are requirements for data caching.

be reviewing each of these cases in this paper, we do want to raise three key questions that arise when considering ethical concerns at the system level.

First, which mechanisms exist for receiving feedback or information regarding possible ethical concerns? Are there methods for library catalog end-users to note that data for specific authors or subjects is problematic? Are there avenues for these end-users or librarians and catalogers to denote which particular instances use data that have ethical concerns?

Second, what affordances does the system provide to handle data which have ethical concerns? Is the system configurable in a way that disables the use of that data for a particular instance or that disables the feature that uses that data? Which points of control over the display or use of the data are present in the system?

Third, are there avenues for providing feedback regarding these problematic instances of data to the data providers themselves?

Defining the answers to these questions before deploying the features that use external data will help clarify how data that pose ethical concerns can be reported, handled, and possibly addressed at the data source itself.

Conclusion

We started the explorations described in this paper to assess whether there were possibilities for using linked and external data in ways that could enhance user discovery in library catalogs. We used a user-centered approach to design, develop, and evaluate mock-ups and working prototypes using linked data. Our user feedback and evaluations show that the use of linked data can effectively contribute to the user experience by providing context about authors and subjects, supporting browsing and navigation to related library resources, and showing suggestions for related searches. While many questions persist, this work represents the foundation of our ongoing efforts to engage users in the development of better discovery environments by incorporating external linked data into the library catalog.

Competing Interests

The authors have no competing interests to declare.

Acknowledgements

Ethics and Consent

Protocols for the user studies in this work were submitted to the Cornell Institutional Review Board (IRB) for Human Participants and determined to be exempt from further review.

References

Allison, DeeAnn. 2010. "Information Portals: The Next Generation Catalog." *Journal of Web Librarianship* 4 (4): 375–89. https://doi.org/10.1080/19322909.2010.507972.

Allison-Cassin, Stacy, Alison Armstrong, Phoebe Ayers, Tom Cramer, Mark Custer, Mairelys Lemus-Rojas, Sally McCallum et al. 2019. "ARL White Paper on Wikidata: Opportunities and Recommendations." Association of Research Libraries, April 18, 2019. https://www.arl.org/wp-content/uploads/2019/04/2019.04.18-ARL-white-paper-on-Wikidata.pdf. Archived at: https://perma.cc/9CEY-598Q.

"Art and Rare Materials (ARM) BIBFRAME Ontology Extension." 2021. https://github.com/Art-and-Rare-Materials-BF-Ext/arm.

Bates, Marcia J. 1989. "The Design of Browsing and Berrypicking Techniques for the Online Search Interface." *Online Review*. https://pages.gseis.ucla.edu/faculty/bates/berrypicking.html. Archived at: https://perma.cc/HW8R-76L5.

Bates, Marcia J. 2007. "What Is Browsing-Really? A Model Drawing from Behavioural Science Research." *Information Research* 12 (4): paper 330. http://informationr.net/ir/12-4/paper330.html. Archived at: https://perma.cc/3Z5A-QUHD.

Boettcher, Emma. 2020. "Where Linked Data Can Impact User Experience: Findings on Discovery Workflows." Presentation at the LD4 Conference on Linked Data in Libraries, July 2020. https://static.sched.com/hosted_files/ld42020/e0/2020LD4_discovery_boettcher.pdf. Archived at: https://perma.cc/K8A2-A9WR.

Cornell University Library. 2020. "Cornell University Implementation of Blacklight." https://github.com/LD4P/blacklight-cornell/tree/ld4p2.

Coyle, Karen. 2016. "Does FRBR Meet FRBR's Objectives?" In *FRBR Before and After: A Look at Our Bibliographic Models*, 97–124. Chicago, IL: ALA Editions. https://kcoyle.net/beforeAndAfter/c8-978-0-8389-1364-2.pdf. Archived at: https://perma.cc/3DGY-LMTF.

Cramer, Tom, Jessie Keck, Huda Khan, Dan Scott, and Simeon Warner. 2019. "A Knowledgeable Panel on Knowledge Panels." Presentation at the LD4 Conference on Linked Data in Libraries, Boston, MA, May 2019. https://docs.google.com/presentation/d/16w9IBxYRmXh7G66vly4y5pR9D0GBjy-j0h6v7bxMOjw/.

Cuna, Andrea, and Gabriele Angeli. 2020. "Improving the Effectiveness of Subject Facets in Library Catalogs and Beyond: A MARC-Based Semiautomated Approach." *Library Hi Tech.* https://doi.org/10.1108/LHT-07-2019-0132.

Dempsey, Lorcan. 2012. "Thirteen Ways of Looking at Libraries, Discovery and the Catalogue: Scale, Workflow, Attention." *Catalogue 2.0: The Future of the Library Catalogue*, edited by Sally Chambers, 179–202.

Diao, Junli. 2018. "Evolution of Western Library Catalogs: The Rising Expectations of Users." *International Journal of Librarianship* 3 (1): 78–93. https://doi.org/10.23974/Ijol.2018.vol3.1.69.

Discogs. 2018. "Discogs API Terms of Use." https://support.discogs.com/hc/en-us/articles/360009334593-API-Terms-of-Use.

Elguindi, Anne C. and Kari Schmidt. 2012. "4 - Discovery systems, layers and tools, and the role of the electronic resources librarian". Editor(s): Anne C. Elguindi, Kari Schmidt. *Chandos Information Professional Series, Electronic Resource Management*, 109-139. ISBN 9781843346685. https://doi.org/10.1016/B978-1-84334-668-5.50004-3.

"Ethics in Linked Data Affinity Group." 2020. https://wiki.lyrasis.org/display/LD4P2/Ethics+in+Linked+Data+Affinity+Group. Last modified September 13, 2021. Archived at: https://perma.cc/XKR3-UW4B.

Fresnido, Ana Maria B., and Agnes S. Barsaga. 2020. "Transaction Log Analysis of OPAC Searches in an Academic Library: Basis for OPAC Interface Improvement." *Philippine Journal of Librarianship and Information Studies* 39 (1): 29–40. https://phjlis.org/index.php/phjlis/article/view/5. Archived at: https://perma.cc/276M-K3MT.

Golub, Koraljka. 2018. "Subject Access in Swedish Discovery Services." *Knowledge Organization* 45 (4): 297–309. https://doi.org/10.5771/0943-7444-2018-4-297.

Gross, Tina, Arlene G. Taylor, and Daniel N. Joudrey. 2015. "Still a Lot to Lose: The Role of Controlled Vocabulary in Keyword Searching." *Cataloging & Classification Quarterly* 53 (1): 1–39. https://doi.org/10.1080/01639374.2014.917447.

Gusenbauer, Michael, and Neal R. Haddaway. 2021. "What Every Researcher Should Know About Searching—Clarified Concepts, Search Advice, and an Agenda to Improve Finding in Academia." *Research Synthesis Methods* 12 (2): 136–47. https://doi.org/10.1002/jrsm.1457.

Hall, Catherine E. 2011. "Facet-Based Library Catalogs: A Survey of the Landscape." *Proceedings of the American Society for Information Science and Technology* 48 (1): 1–8. https://doi.org/10.1002/meet.2011.14504801195.

International Federation of Library Associations and Institutions (IFLA) Working Group on the Functional Requirements for Subject Authority Records (FRSAR). 2010. *Functional Requirements for Subject Authority Data (FRSAD): A Conceptual Model.* https://www.ifla.org/wp-content/uploads/2019/05/assets/classification-and-indexing/functional-requirements-for-subject-authority-data/frsad-model.pdf. Archived at: https://perma.cc/9HJU-BJFC.

Intarapaiboon, Peerasak, and Chainarong Kesamoon. 2019. "Applying Domain Knowledge and Academic Information to Enhance Unknown-Item Search in OPAC." *Malaysian Journal of Library & Information Science* 24 (1): 45–58. https://doi.org/10.22452/mjlis.vol24no1.3.

Johnson, Thomas, and Karen Estlund. 2014. "Recipes for Enhancing Digital Collections with Linked Data." *Code4Lib Journal* 23. https://journal.code4lib.org/articles/9214. Archived at: https://perma.cc/LJZ5-A865.

Jin, Qiang, James F. Hahn, and Gretchen Croll. 2016. "BIBFRAME Transformation for Enhanced Discovery." *Library Resources and Technical Services* 60 (4). http://hdl.handle.net/2142/90248.

Khan, Huda. 2021. "Discovery on the Ground". https://wiki.lyrasis.org/display/LD4P3/Discovery+On+the+Ground. Archived at: https://perma.cc/GPY5-HJ9W.

Khan, Huda, and Steven Folsom. 2021. "Production:Discogs." https://wiki.lyrasis.org/display/LD4P3/Production%3A+Discogs. Archived at: https://perma.cc/H8GJ-53NL.

Khan, Huda, Tim Worrall, and John Skiles Skinner. 2020a. "BAM! Lessons Learned and Potential Future Work." https://wiki.lyrasis.org/display/LD4P2/BAM%21+Lessons+Learned+and+Potential+Future+Work. Archived at: https://perma.cc/TPV7-TAYB.

Khan, Huda, Tim Worrall, and John Skiles Skinner. 2020b. "SMASH! Lessons Learned and Future Work." https://wiki.lyrasis.org/display/LD4P2/SMASH%21+Lessons+learned+and+future+work. Archived at: https://perma.cc/WZ74-FY9M.

Khan, Huda, Tim Worrall, and John Skiles Skinner. 2020c. "WHAM! Lessons Learned." https://wiki.lyrasis.org/display/LD4P2/WHAM%21+Lessons+Learned. Archived at: https://perma.cc/9QFK-9KZY.

Khan, Huda, Tim Worrall, and John Skiles Skinner. 2019. "KPAOW! Lessons Learned and Possible Future Steps." https://wiki.lyrasis.org/display/LD4P2/KPAOW%21+Lessons+Learned+and+Possible+Future+Steps. Archived at: https://perma.cc/9EJQ-2T8X.

Kovari, Jason, Steven Folsom, and Rebecca Younes. 2017. "Towards a BIBFRAME Implementation: The Bibliotek-o Framework." Paper presented at the DCMI International Conference on Dublin Core and Metadata Applications, Washington, DC, October 2017. Archived at: https://perma.cc/DV2W-4HP9.

Kungliga biblioteket. n.d. "LIBRIS." Accessed June 25, 2021. https://libris.kb.se/.

Library of Congress. n.d. "BIBFRAME Ontology." Accessed June 25, 2021. https://id.loc.gov/ontologies/bibframe.html. Archived at: https://perma.cc/KY48-E5AZ.

Library of Congress. n.d. Bibliographic Framework Initiative. Accessed November 23, 2021. https://www.loc.gov/bibframe/. Archived at: https://perma.cc/Z2UB-36QL.

Library of Congress. 2020. "MARC Standards." Last modified March 13, 2020. https://www.loc.gov/marc/. Archived at: https://perma.cc/545V-SLNE.

Mak, Lucas, Devin Higgins, Lisa Lorenzo, and Sruthin Gaddam. 2021. "What Am I Looking at: Contextualizing Subject Headings Through Linked Open Data." *College & Undergraduate Libraries* 27 (2–4): 435–51. https://doi.org/10.1080/10691316.2021.1901634.

Martin, Jennifer M. 2021. "Records, Responsibility, and Power: An Overview of Cataloging Ethics." *Cataloging & Classification Quarterly* 59 (2–3): 281–304. https://doi.org/10.1080/01639374.2020.1871458.

Martin, Kristin E., and Kavita Mundle. 2010. "Cataloging E-Books and Vendor Records." *Library Resources & Technical Services* 54 (4). https://doi.org/10.5860/lrts.54n4.227.

McGrath, Kelley. 2007. "Facet-Based Search and Navigation with LCSH: Problems and Opportunities." *Code4Lib Journal* 1. https://journal.code4lib.org/articles/23. Archived at: https://perma.cc/R25D-GKWU.

McKay, Dana, George Buchanan, and Shanton Chang. 2018. "It Ain't What You Do, It's the Way That You Do It: Design Guidelines to Better Support Online Browsing." *Proceedings of the Association for Information Science and Technology* 55 (1): 347–56. https://doi.org/10.1002/pra2.2018.14505501038.

McKay, Dana, Shanton Chang, Wally Smith, and George Buchanan. 2019. "The Things We Talk About When We Talk About Browsing: An Empirical Typology of Library Browsing Behavior." *Journal of the Association for Information Science and Technology* 70 (12): 1383–94. https://doi.org/10.1002/asi.24200.

Meyer, Steve. 2016. "Enhancing Library Discovery with Linked Open Data." https://web.library.wisc.edu/sp/eluna-igelu-showandtell/. Archived at: https://perma.cc/6QRN-X9ZT.

Naun, Chew Chiat. 2020. "Expanding the Use of Linked Data Value Vocabularies in PCC Cataloging." *Cataloging & Classification Quarterly* 58 (3–4): 449–57. http://doi.org/10.1080/01639374.2019.1705953.

Palagi, Emilie, Fabien Gandon, Alain Giboin, and Raphaël Troncy. 2017. "A Survey of Definitions and Models of Exploratory Search." In *ESIDA '17: Proceedings of the 2017 ACM Workshop on Exploratory Search and Interactive Data Analytics*, 3–8. https://doi.org/10.1145/3038462.3038465.

Riva, Pat, Patrick Le Bœuf, and Maja Žumer. 2017. *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. International Federation of Library Associations and Institutions. https://repository.ifla.org/handle/123456789/40.

Schilling, Virginia. 2012. "Transforming library metadata into linked library data." *Association for Library Collections and Technical Services*. http://www.ala.org/alcts/resources/org/cat/research/linked-data. Accessed April 4, 2022

Snyder, Eunah, Lisa Lorenzo, and Lucas Mak. 2019. "Linked Open Data for Subject Discovery: Assessing the Alignment Between Library of Congress Vocabularies and Wikidata." Paper presented at the DCMI International Conference on Dublin Core and Metadata Applications, Seoul, September 2019. https://dcpapers.dublincore.org/pubs/article/view/4225/0.html. Archived at: https://perma.cc/69N7-9MH8.

Stahmer, Carl G. 2020. "A World of Curated Knowledge: Leveraging the Wider Semantic Web to Enhance Library Discovery." In *Resource Discovery for the Twenty-First Century Library: Case Studies and Perspectives on the Role of IT in User Engagement and Empowerment*, edited by Simon McLeish, 147–60. London: Facet. https://doi.org/10.29085/9781783301409.

Stanford University Libraries. 2020. "SearchWorks." https://github.com/sul-dlss/SearchWorks/tree/linked-data-experiments.

Sullivan, Daniel. 2020. "A Reintroduction to Our Knowledge Graph and Knowledge Panels." The Keyword. https://blog.google/products/search/about-knowledge-graph-and-knowledge-panels/. Archived at: https://perma.cc/88FB-CXQJ.

Trapido, Irina. 2016. "Library Discovery Products: Discovering User Expectations Through Failure Analysis." *Information Technology and Libraries* 35 (3): 9–26. https://doi.org/10.6017/ital.v35i3.9190.

Usong, Astrid. 2019. "Discovery User Research." https://wiki.lyrasis.org/display/LD4P2/Discovery+user+research. Archived at: https://perma.cc/3PNG-65T9.

W3C. 2014. "Resource Description Framework (RDF)." https://www.w3.org/RDF/. Archived at: https://perma.cc/B329-KHT6.

Wakeling, Simon, Paul Clough, Lynn Silipigni Connaway, Barbara Sen, and David Tomás. 2017. "Users and Uses of a Global Union Catalog: A Mixed-Methods Study of WorldCat.org." *Journal of the Association for Information Science and Technology* 68 (9): 2166–81. https://doi.org/10.1002/asi.23708.

Wakimoto, Jina Choi. 2009. "Scope of the Library Catalog in Times of Transition." *Cataloging & Classification Quarterly* 47 (5): 409–26. https://doi.org/10.1080/01639370902865371.

Wells, David. 2020. "Online Public Access Catalogues and Library Discovery Systems." *ISKO Encyclopedia of Knowledge Organization.* Edited by Birger Hjørland and Claudio Gnoli. https://www.isko.org/cyclo/opac. Archived at: https://perma.cc/ZJ5H-RLJS.

Wisser, Katherine M. 2014. "The Errors of Our Ways: Using Metadata Quality Research to Understand Common Error Patterns in the Application of Name Headings." In *Metadata and Semantics Research: 8th Research Conference,* MTSR 2014, Karlsruhe, Germany, November 27–29, 2014, Proceedings, edited by Sissi Closs, Rudi Studer, Emmanouel Garoufallou, and Miguel-Angel Sicilia, 83–94. Cham: Springer. https://doi.org/10.1007/978-3-319-13674-5_9.

Yasser, Chuttur M. 2011. "An Analysis of Problems in Metadata Records." *Journal of Library Metadata* 11 (2): 51–62. https://doi.org/10.1080/19386389.2011.570654.

OPEN ACCESS