

KPAOW! Lessons Learned and Possible Future Steps

[Original Google Docs link](#)

User research and usability overview

High level results from earlier evaluations as well as mockup evaluations we did at Cornell indicate knowledge panels that represent contextual information as well as related relationships that can lead to further searches in the catalog can be useful. People's perspectives on which pieces of information are deemed relevant for context and for additional links to search appear appeared to differ. 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 KPAOW mockup evaluation participants knew what content was entailed by digital collections. Future work could focus on differences between knowledge panel layouts (e.g within page or 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.

More in-depth information about the review of articles, user interviews, and usability results can be found here: <https://wiki.duraspace.org/pages/viewpage.action?pageId=101783940>.

KPAOW mockup evaluation results are available here: https://docs.google.com/document/d/1BFWtki1LfmHsTtz6C25M9ZNv1GmJTN_VVg-fa34E6Ys/edit.

Data sources, APIs, and analyses

The data sources selected and used for retrieving data or making connections for KPAOW were LCNAF, LCSH/FAST, DbPedia, Wikidata, and Discogs. Additionally, we set up a proxy controller within the Blacklight app to directly access the Digital Collections Blacklight catalog search results as JSON. We also looked at using OCLC Work Ids from the catalog to find the appropriate Wikidata Ids. Details can be found in the development-related results below. A quick summary of data sources and where they were used is found in the table below.

Source	Used for	Overview of properties and queries employed
Id.loc.gov (LCNAF/ LCSH)	<p><i>Knowledge Panel</i></p> <p>Retrieving URI/local name based on authority string</p> <p><i>Search Suggestions</i></p> <p>Retrieving terms to populate suggested searches list on catalog search page based on query</p>	<p>"http://id.loc.gov/authorities/[auth]/suggest/?q=" where [auth] is either "names" or "subject"</p>
Wikidata	<p><i>Knowledge Panel</i></p> <p>Retrieving Wikidata entity based on LC local name</p> <p>Retrieving image URL, "notable works", "influenced", "influenced by", and "awards received" properties for a Wikidata entity</p> <p><i>Related works</i></p> <p>Retrieving derived works and editions for the work entity identified by OCLC Work ID in the catalog</p> <p><i>Search Suggestions</i></p> <p>Retrieving terms to populate suggested searches list on catalog search page based on query</p> <p><i>Item view</i></p> <p>For Discogs examples, the Wikidata entity related to the work was also retrieved and used to retrieve the previous and following works for a particular item</p>	<p>Use wdt:P244 to retrieve Wikidata URI based on LC local name</p> <p>Properties wdt:P800 (notable work), wdt:P737 and wdt:P737 (influence relationships), wdt:P18 (image)</p> <p>Properties wdt:P629 (editions) and wdt:P4969 (derivatives)</p> <p>Previous/following works wdt:P155 and wdt:P156</p> <p>Discogs master ID</p> <p>wdt:P1954</p>
DbPedia	<p><i>Search Suggestions</i></p> <p>Retrieving terms to populate suggested searches list on catalog search page based on query</p>	
Discogs	<p><i>Item view</i></p> <p>Retrieving Discogs properties to include within the item view for a particular search result. Discogs was searched based on item label/title.</p>	

Digital Collections	<p><i>Knowledge Panel</i></p> <p>Retrieving digital collection results using a keyword search to Digital Collections Blacklight instance using the authorized heading for name or subject heading for the knowledge panel</p> <p><i>Subject heading facet</i></p> <p>Retrieving Digital Collections subject results using the subject heading selected in the faceted search.</p>	Subject heading facet search uses "subject_tesim" as the Digital Collections facet
---------------------	---	--

OCLC Work Ids appear to have changed format from what is captured in Wikidata and what are the most current versions. Having updated IDs in Wikidata and/or the catalog as well as having more OCLC work id to Wikidata connections would enable for greater linkages and the possibility for bringing back more information.

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. The code had to execute queries that first retrieved the LOC entity local name (i.e the last part of the URI) by searching against id.loc.gov and then conducting 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 result of this search was then used to match against possible URIs in Wikidata. Having the appropriate URIs present in the search index would enable us to bypass these two queries.

Development-related results

- Discogs Data

Successfully supplemented the catalog metadata for some musical recordings by bringing in information from the Discogs database. All ajax calls to the Discogs API went through the Authorities Lookup. Those initial search queries, however, had to rely on keywords rather than IDs because the catalog index does not contain any Discogs IDs or URIs. As a result, the criteria established to prevent mismatches between a work in the catalog and a Discogs record would result in the rejection of what were actually valid matches between the catalog and Discogs. Similarly, some slight differences between the catalog and Discogs metadata also prevented pairing up a work in the catalog with the corresponding work in Discogs.

- Work-level Wikidata

Using the OCLC work IDs in the catalog index, for a given work we linked to Wikidata to display lists of both derivative works (for example, film adaptations of a work) and significant editions of the work. In a couple of cases, the OCLC work ID that the catalog contains for a work (such as The Adventures of Tom Sawyer) did not correspond to the work that Wikidata has for the same OCLC work ID (The Adventures of Huckleberry Finn). More significantly, only a small percentage of the items in the catalog -- ~ 0.2% -- have matches in Wikidata..

- Related Searches Display

Modified the catalog's main search feature so that we take the search term and query it against Wikidata, DBpedia and the Library of Congress subject authority. The results of those queries are combined into a single list of terms, with duplicates removed, and then we search each term against the catalog. If there's a match to the catalog, the list of terms is then displayed to the user as possible related or "alternative" searches. There are two primary drawbacks here: first, the time it takes to do these queries is excessive; and second, the results can contain terms or subjects that are not actually related to the initial search term.

- Knowledge Panel

Added a "popover" box that can be activated from an icon appearing next to author and subject lines in item record view. The box provides wikidata information when available, including birth and death dates and locations, occupation, and an image. If the entity is an author, notable works and influences by and for that person appear in the knowledge panel. Finally, the knowledge panel provides links into Cornell's Digital Collections when related results or contributors can be found.

Possible development enhancements

The objective for this set of code was to serve as a proof of concept or demo for the integration of linked data in the library catalog. Our focus was on functionality. Much can be done with respect to refactoring both the code and the employed SPARQL queries as well as adding tests to help make this code production ready. That said, the code can serve as a starting point for future enhancements and/or further experiments.

Possible future work could include:

- Handling any remaining bugs or known areas for reviewing code, such as for event handling around knowledge panel triggering and closing
- Adding tests for front-end JavaScript code as well as extensions for partials displaying subject area facets
- Integrating Digital Collections results for author facet selections
- Passing along keyword searches within the catalog to Digital Collections as well, thereby perhaps enabling a keyword and facet combined query, i.e. to pass along the entirety of the query in the catalog to Digital Collections
- Combining the queries to Wikidata into a single query while making sure to test for performance/optimization of the queries in question
- Adding LOC and Wikidata identifiers to the catalog or an external search index to allow for quick retrieval of this information without having to execute AJAX queries to first obtain a possible LOC local name and then its corresponding Wikidata identifiers. VIAF could also be considered as a possible hub for retrieving appropriate information (See University of Wisconsin's work in this regard).
- Knowledge panels could further integrate "works by" and "works about" sections into Digital Collections results.
- Additional work could be done in the areas of UX design work and usability testing against the prototypes themselves to provide feedback on design possibilities
 - UX work and design around integration of "browse by subject" for the item view would also be a potential area for future work

- Subject subdivisions are broken out by hierarchy using indentations. We should review better interface design options for representing this breakout or hierarchies in the UI.
- Knowledge panels contain a substantial amount of information and reviewing usability for design and included information would be beneficial

Work to extend the browse by subject implementation will be explored in the next work phase (BAM!).