Objectives for Digital Collections in Shared Primo

Orbis Cascade Alliance

Unique & Local Content Team, Digital Collections in Primo Project Group

2018 June 14

Project Group Members

> Ann Lally, University of Washington, Chair
> Sara Amato, Willamette University
> Philip Vue, University of Portland
> Dan Oswalt, Lewis & Clark College
> Jodi Allison-Bunnell, Alliance (ex officio)

Executive Summary

The [Digital Collections in Primo Project Group](#) (DCP PG) continued the work of the [Digital Collections Working Group](#) (DC WG) to create objectives for digital collections in the Alliance Primo installation. It is essential that the Alliance has a shared understanding of what success looks like in order to determine whether implementation is desirable or feasible.

The objectives are based on the original purpose of this project as approved by Council in FY16, the outcomes of FY17 and FY18 user testing, and implementing a pipe from the digital collections harvester to the Alliance's Premium Sandbox in 2018 and doing base-level work on normalization rules and Primo Back Office settings to enable user testing. They were presented and discussed with the ULC community in an open forum in February.

We propose the following objectives for including digital collections successfully in shared Primo:
- Enable discovery of all materials held by Alliance libraries in one space.
- Centralize digital collections aggregation through the Digital Collections Harvester so that members can stop piping things individually if they choose.
- Include digital collections in the general search scope by default rather than breaking them out in a separate category, but allow institutions to opt in or out according to their needs and preferences.
- Ensure that the pathway to the digital object is clear in both search results and the item record. Specifically, provide a means for access from both text and visual links.

- Incorporate nested facets for Resource Type. Card sort testing suggested top level facets of Images, Maps, Text, Video, Audio, and Software, with more specific terms nested in those top-level categories.
- Provide standardized and free-text rights statements, and make the standardized rights statements available as a facet.
- Consider the identity of the institution as secondary. End users do not care which institution holds a particular object if they can access it online unless they need to pay or request permissions. However, the institutions do consider this an important matter, and that needs to be accommodated.

Background and Purpose of the Objectives

In March 2016, Alliance Council approved funding to support digital collections at the Alliance by building a harvester for digital collections, support metadata remediation, and apply to become a DPLA hub for Alliance members. That initiative has since become the Digital Collections Service of the Unique & Local Content Program, and the Alliance became a DPLA hub in Spring 2018. Further background is available here.

As part of the approval process, Council specified that the Alliance should include making digital collections discoverable in Alliance Primo, pending assessment of whether doing so was successful. Support for this addition was also based on the efficiency of maintaining a single pipe from the harvester into Primo (rather than each institution maintaining one or more pipes), and maximizing the use of an infrastructure in which we have heavy investment.

Subsequent to this approval, many individuals expressed doubts about the viability of this portion of the project, asking questions like "Is Primo ready for digital collections?" and "Won't users just be overwhelmed by thousands of digital objects?" In order to answer those questions, Alliance members need a shared sense of what "ready" means, and what uses cases we want to support with this function. This question poses particular challenges as institutions expand their thinking about digital collections from a local presentation to presentation in large aggregation.

Since then, the Alliance has twice received LSTA funding to clean up digital collections metadata at Alliance institutions. We have developed the digital collections harvester that produces output specifically normalized for Primo. The Digital Collections Working Group conducted user testing with end users. As of this writing, the harvester has about 40,000 digital objects in it for Primo, and there is a pipe into the Alliance's Premium Sandbox for testing. The Shared and Unique Metadata in Discovery ad hoc group worked through a list of potential issues and solutions as their first pilot project. The Normalization Rules Standing Group has continued work on resolving issues with search, display, and function in Primo.

In charging the Digital Collections in Primo Project Group, the ULC Team asked that the objectives include:

- Overall intent of this initiative (which is almost certainly not widely understood)
- How users get to the digital object in the source repository
- Institution identity: End users are not particularly interested in this, but institutions are.
- Facets: How do we effectively label digital objects from DAMs or IRs in a search interface that includes all manner of library materials? Do these items need a special category, or are they more usefully integrated with published materials?
- Discovery by institution vs. all institutions. What is the default option? What options do we offer institutions?

Digital objects and collections have been defined by the Unique & Local Content Team as:
- Held by only a single library and cannot be readily replaced or substituted;
- Held in a very small number of libraries, in aggregations unique to the institution; or are original works initially published/distributed from within the institution.

Thus, "digital collections" are made up of born-digital or digitized objects, most commonly:
- Faculty and student scholarly and creative work (e.g. what you often find in an IR)
- Images (still/moving); audio, archives, records, and manuscripts; college and university publications (e.g. what you often find in a DAM)

Background Knowledge and Assumptions

The group worked with the following knowledge and assumptions:
- Undergraduates start with search engines rather than library catalogs to begin their research;
- Primo does not have search engine exposure at present;
- OCLC does not have data regarding the success of making large numbers of digital objects available in WorldCat now for 10+ years;
- Most researchers find unique materials using search engines; a few will use specific sites as their starting point.

Further, the group understood Council's objectives for this initiative thus:
- Enable discovery of all materials held by Alliance libraries in one space
  - Ensure clear designation for items to be viewed online
- Centralize process so that local organizations can stop piping things individually
  - Provide mechanisms for local organizations to opt in based on need

Work of Digital Collections Working Group (FY17)

Previous work by the Digital Collections Working Group during 2016-2017 informed our work. Their report is [available on the Alliance website](#), but a summary of their findings from their user testing is:

- Provide clear and consistent pathways from Primo to the source repository (CONTENTdm, Digital Commons, etc.) that utilize both text and visual cues.
- Institution identity (e.g. which institution holds the original) only matters to end users if they need to request access, pay, or seek permissions. Acknowledge institutions' need to attach their identity to their signature collections, but do so in a way that does not affect usability.

The group also tested Resource Type facets in comparator Primo sites (Brigham Young University, Royal Danish Library, University of Tennessee), but was unable to come to any conclusions. So, as part of their final report the Digital Collections Working Group recommended that future work include "a card sorting test to understand how users organize facets and what language/terms are intuitive to them."

The group was unable to complete work on objectives in the time it had, so that work plus a Premium Sandbox implementation was passed to the FY18 group.

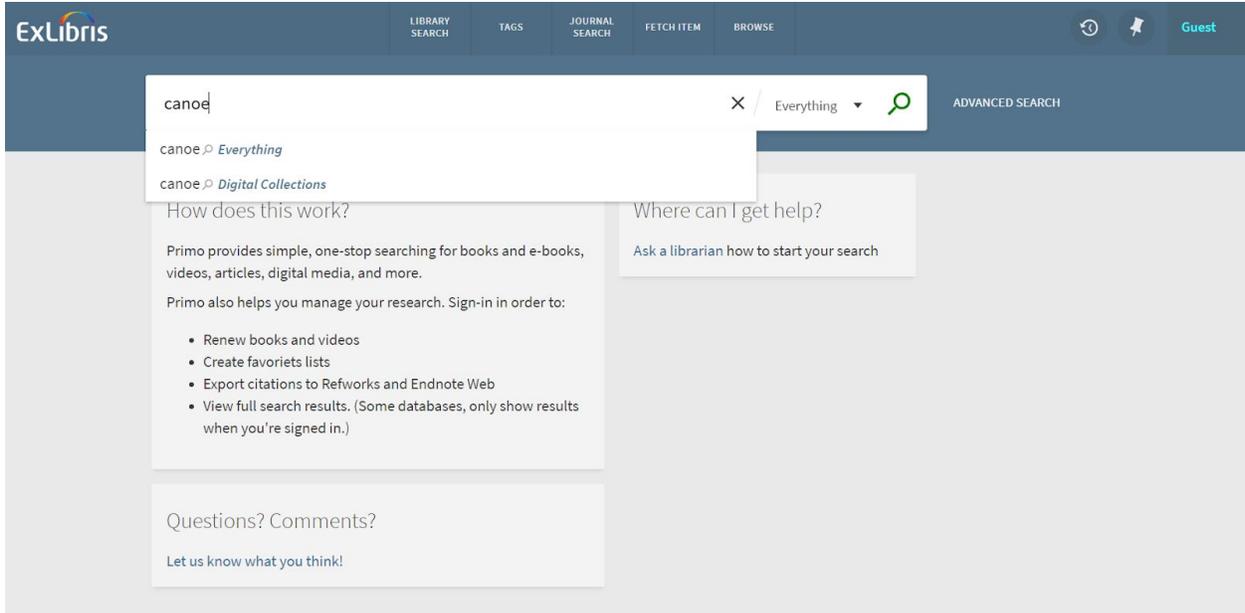# Work of Digital Collections in Primo Project Group (FY18)

Topics of Focus

Based on the recommendations of the Digital Collections Working Group after their user testing, this group focused on two topics in user testing:
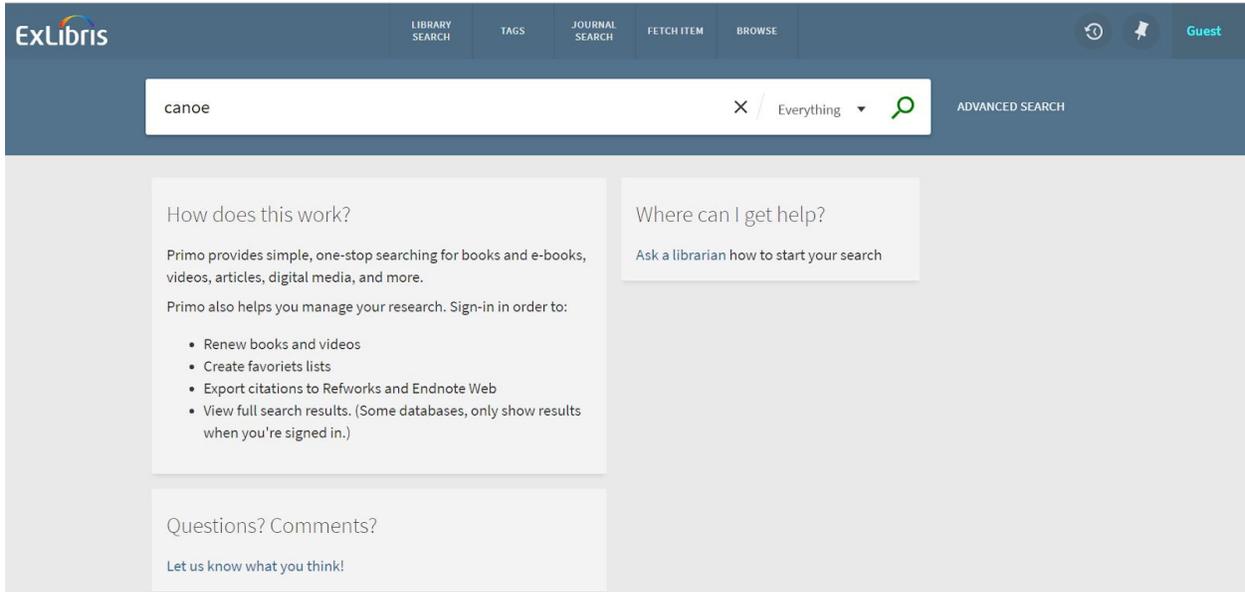- Top search box
- Type and Genre Facets

Recommendation: Top Search Box

Determine if digital objects go into a special search scope or along with other materials or both. Do we want digital objects to behave the same way or differently because they are special?

User testing results say: users do not want to narrow their search at this phase; they want to see all results, then choose to narrow at that point.

We recommend that the top search box include all digital objects by default, as pictured below.

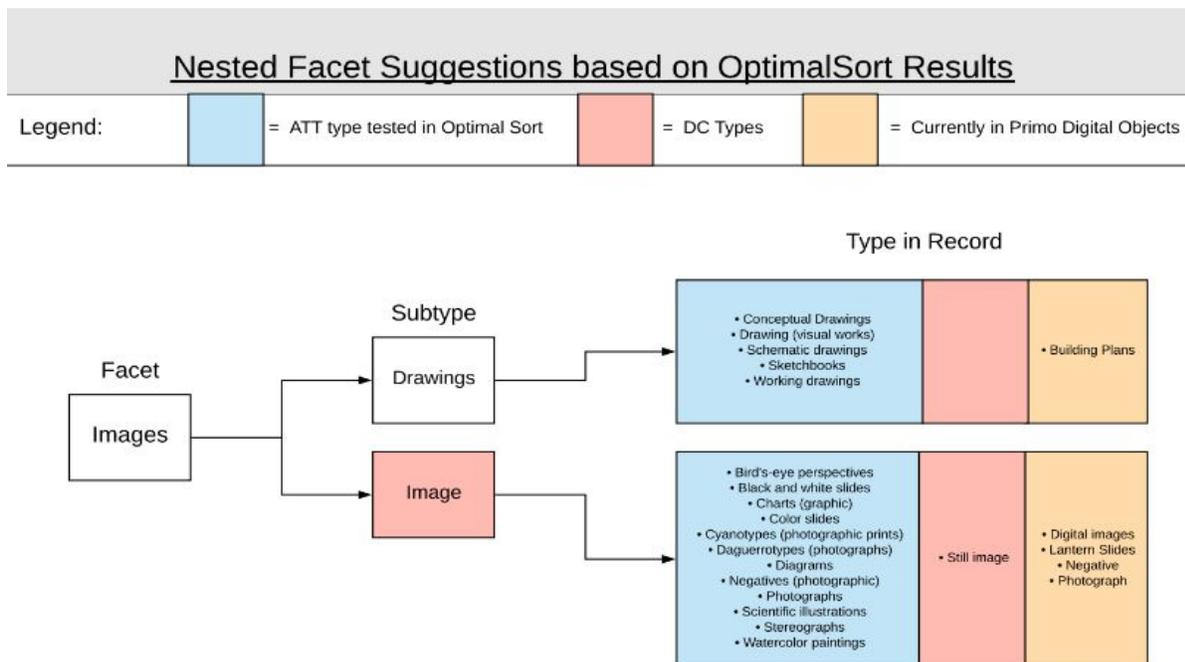## Recommend Facets for Type and Genre
How do we effectively categorize and label digital objects from DAMs or IRs in a search interface that includes all manner of library materials? E.g. A box of print images available in an archives (metadata only available online) vs. a digitized image available online -

both could be in a resource type facet of images, but they are distinct in their modes of access. Should digital objects go into a category under a 'type/genre' facet or be in a facet by themselves?   What is the relationship between facets for 'Resource Type', 'Availability' and 'Physical Location'?

We are assuming we will be exploring both facet labels and types of materials that would fall into those labels.

User testing results say: facets are liked and used by those conducting searches, particularly nested facets. Nested facets should be implemented under Resource Type rather than Physical/Online Availability. Test subjects reported that the resource type is more meaningful to them than availability of unique versus commodity content. Further details can be found in the User Testing Report.

During the large card sort users were asked to group like items into broader categories. Test results show these broad categories were similar enough to suggest potentially useful broad facet terms of Images, Maps, Text, Video, Audio, Software, under which more specific terms could be nested. The illustration below suggests the relationship between the top-level and narrower facets for images.

The illustrations below suggest how this could work in Primo by showing the present, inspiration from a comparator site, and a possible ideal.

Now: Alliance Primo Premium Sandbox with harvester content piped in	Inspiration: Brigham Young University Primo instance	Implementation Ideal: Nested facets for both Alma and harvester content
Resource Type ^ 📌 ↑	**Type** ^	**Resource Type** ^ 📌 ↑
Physical Books *(144,059)*	☐ Articles (133,286)	Audio
Government Docu... *(68,393)*	☐ Books (7,126)	eAudio
Dissertations *(62,442)*	☐ E-books (4,046)	Audiocassettes
eBooks *(27,034)*	☐ Videos (61)	Audio CDs
Journals *(15,708)*	☐ Blu-rays (1)	Images
Maps *(13,782)*	☐ DVDs (23)	Drawings
Images *(8,454)*	HIDE ↑	Photographs
Text Resources *(6,885)*	☐ Filmstrips (1)	Sketchbooks
Microform *(6,438)*	☐ VHS (6)	Maps
Archival Materials *(4,689)*	☐ Audio (155)	Software
Computer files *(3,464)*	☐ Audiobooks (63)	Computer files
Videocassettes *(1,930)*	☐ CDs (18)	CD-ROM/DVD-ROM
Newspapers *(1,787)*	☐ Musical Recordings (30)	Text
Audio CDs *(1,370)*	☐ Records (17)	Books
DVD Videos *(1,054)*	☐ Conference Proceedings (4)	Dissertations
eAudio & eVideo *(695)*	☐ Databases (2)	eBooks
Audiocassettes *(585)*	☐ Datasets (1)	Government documents
Scores *(569)*	☐ Electronic Resources (3)	Journals
CD-ROM/DVD-ROM *(434)*	☐ Images (1)	Manuscripts
Loose-Leaf *(419)*	☐ Manuscript Archives (7)	Newspapers
Web Sites *(297)*	☐ Microformats (7)	Video
	☐ Music scores (16)	DVD videos
	☐ Other (4)	eVideo
	☐ Periodicals (1)	Videocassettes

[The complete user testing results from FY18](#) are available on the Alliance's website.