UNIVERSITY OF ALBERTA
LIBRARIES

DISCOVERY STRATEGY REVIEW TEAM REPORT

Table of Contents

Background

- The Discovery Strategy Review Team (DSRT) met from December 2018 to August 2019 to study UAL's current discovery environment and recommend changes as appropriate
- At this stage DSRT is not able to recommend a definitive technological solution for discovery at the UAL; however, a Discovery Phase II Concept Design and Preliminary List of Requirements are provided in this report as possible ways forward, pending additional research, that provide UAL with the opportunity to address the findability of our rich digital and physical collections and the sustainability and streamlined operation and development of library discovery systems
- UAL should continue to find ways to make our collections discoverable in as consistent and coordinated a manner as possible without sacrificing comprehensiveness or over-simplifying search functionality
- Ongoing assessment of open source and vendor-based discovery solutions, and the appropriate balance of each in UAL's discovery environment is required, and should be further assessed prior to any additional major investments in discovery development

Submitted by the Discovery Strategy Review Team (Sean Luyk, Christopher Bateman, Ian Bigelow, Trish Chatterley, Sharon Farnel, Weiwei Shi, Sandra Shores, David Sulz, Leah Vanderjagt)

2019-08-12

Recommendations

Based on an analysis of the current discovery environment, including information gathered both internally and externally, DSRT submits the following recommendations for consideration:

1. UAL commit to forming a Discovery Operations Team (DOT), separate from the Web Architecture Team (WAT), responsible for the ongoing maintenance and development of discovery (draft provided in Operational Approach). DSRT determined that discovery would benefit from more regular and focused attention than is possible within WAT's current structure and mandate. Its activities will include: gathering and developing user requirements, soliciting and responding to user and staff feedback, developing use cases and prototypes, software testing/quality assurance, and liaising with the UAL staff (dev, sys admin, bib services team members) assigned to discovery
    1.1. The first priority of this group will be to assess the information provided by discovery vendors during DSRT's investigation, and to determine what role they should play in the future of UAL's discovery environment and its relationship to open source development of discovery. DSRT recommends that this assessment be completed by the end of 2019
    1.2. DOT should also commit to working with partner institutions, initiatives, and collaborations to further develop options for our discovery ecosystem
2. Dependent on available resources, and in light of the information gathered in recommendation 1.1, UAL begin developing Discovery Phase II, a set of improvements to discovery focusing on user satisfaction and operational sustainability. Discovery Phase II consists of:
    2.1. Adoption of a "coordinated bucket approach" to the UI and underlying infrastructure (see Phase II Concept Design and Technical Approach for more details)
    2.2. A UI redesign focused on reducing complexity, improving visual appeal and usability, and providing a more accessible user experience (see UI Mockups)
    2.3. Creation of a local collections view that enables discovery of collections that are currently siloed
    2.4. Re-introduction of an articles view that presents results from EDS or another knowledge base
    2.5. Improvements to search relevance, and the development of features to support precision searching and browsing
    2.6. Ongoing commitment to maintaining discovery, including reducing technical debt, simplifying discovery applications and associated infrastructure to ensure they remain manageable and sustainable, improving test coverage to ensure that changes to the applications yield expected results, and high priority bug fixes
3. UAL commit to hiring or redeploying a Data Manipulation Specialist (1 FTE) to support the ongoing ingest of new collections to discovery, development of data transformation scripts, and the configuration/development of data warehousing tools. DSRT determined that this was one area that was lacking in the current staff complement for discovery. The successful candidate for this position would, first and foremost, require expertise in metadata, followed by knowledge of data programming techniques

Methodology

- The DSRT utilized a business analysis approach to developing strategy that involved documenting and analyzing the current state of discovery and developing a future state recommendation based on findings from this analysis
- We looked at four aspects of discovery at the UAL: people, processes, information, and technology, and conducted documentation activities in each category for both the current and future state
- The scope of our analysis was mainly focused on the internal environment, but this was balanced by conducting an environmental scan and reviewing current literature on library discovery trends

Current State Analysis

Problem Overview

The people, processes, information, and technology that comprise library discovery at the UAL require a renewed strategy that is operationally sustainable and prepared to serve the evolving needs of library users. Library discovery, broadly defined, impacts the ability of users to access and use the wealth of resources contained in UAL's collections in support of their learning, teaching, and research activities. For the UAL, library discovery represents a crucial network of technologies and services that allow the organization to achieve its 2019-2020 strategic priorities, most notably as they relate to the curation and management of digital assets, our focus on accessibility and usability, support for signature teaching/research areas, and our participation in the transition to linked open data.

The Discovery Strategy Recommendation (2013) paved the way for a discovery strategy that focused on improving the user experience, bringing disparate information silos (SFX, databases) closer together by leveraging the work of the open source library software community (Blacklight, Solr), and moving us away from vendor-based discovery solutions. This strategy proved to be beneficial to library users, but also challenging to sustain from an operational/technical perspective, as Blacklight has proven to be difficult to extend to serve broader purposes for discovery such as indexing items contained in digital collections and repositories, and displaying them in meaningful ways to users. Given the investment UAL has made in developing its own discovery solution and growing expertise in Solr, it is important going forward to build on and take advantage of expertise gained, but also to continuously look for other technologies that may help us to meet emerging, innovative needs of discovery by our users.

External Environment

Discovery Vendor Landscape

The current discovery landscape is dominated by three vendors: ProQuest/ExLibris (Summon, Primo), EBSCO (EBSCO Discovery Service), and OCLC (Worldcat Local). Each of these vendors offer "all in one" solutions that provide capabilities for metadata ingest, indexing, and search through a standardized UI that allows for some level of customization. The Discovery Strategy Recommendation (2013) noted that "given their lack of flexibility and customizability, as well as their cost and difficulty of installation and maintenance, none of the commercial discovery solutions are adequate on their own for a library-wide discovery strategy." Vendor and locally developed systems both come with significant costs in different ways, but compared to a UAL developed discovery solution, vendor-based solutions alone tend to provide less flexibility and customizability. As many past experiences in collections and IT have also shown us, there is a significant risk of vendor lock-in[1] anytime a library moves a large portion of its business to a single vendor. This problem is compounded in the library discovery sector due to the fact that few viable substitutes to the main vendors exist, resulting in high switching costs, in addition to subscription renewal increases known to average around 3-5% per year. Through the course of our future state analysis one finding of the group was that we can expect significant developments in metadata and its application to new forms of discovery in coming years. The requirement for flexibility and custom development suggests that unless a vendor solution can be identified that is closely aligned with identified needs, we can reasonably expect that this would be a less than ideal approach.

In light of these issues, the DSRT contacted Ex Libris, OCLC, EBSCO, Casalini Libri, @Cult, iii, and SirsiDynix to determine if our preliminary list of requirements (see Appendix 1) could be met, and if the issues with vendor-based solutions described in the 2013 Discovery Strategy Recommendation remained true. Vendors were asked to respond to the following questions:

1. Could you please provide us with some example institutions using your discovery system to present local collections such as locally digitized collections and/or institutional repository? How much custom development work (time/resources) is involved to include these local collections?
2. When will your discovery systems support linked data? Outline your approach, timeline and any current test or production implementations. Will we be able to leverage metadata with URIs, or expose that data through SPARQL endpoints? When will you support the use of BIBFRAME data? And At what level?
3. What are the typical API response times? Limitations with your API (rate limit per second/per day/per instance etc)?
4. What level of flexibility (eg. customize metadata fields that are indexed/displayed/faceted, customized thumbnail image display, relevance ranking and sorting, etc.) we have around customization? What is the workflow for such customization? How much custom development work is required to do these customizations? What are the trade-offs and limitations? Can you

[1] See "Vendor Lock-in Definition", The Linux Information Project. <http://www.linfo.org/vendor_lockin.html>

provide us with example institutions where customization has already been incorporated? If not currently possible, what are your timelines for launching this functionality?

5. What aspects of the EDS article KB are open? How can we interact with the article KB, other than through API requests?

Based on our requirements and vendor responses, EBSCO, OCLC, and SirsiDynix were eliminated from consideration (see Appendix 2). Ex Libris showed considerable promise with respect to many of the requirements outlined by the group, but further investigation would be required to determine how specific goals could be met and how flexibility could be built in to address the needs of UAL. Further analysis of Ex Libris can be part of the initial work of the DOT (see Recommendation #1), keeping in mind UAL's previous experience attempting to implement Primo. In 2011, UAL purchased Ex Libris' Primo system and licensed access to the Primo Central knowledge base. In March 2013, the implementation of Primo was cancelled due to unresolvable issues with ingest of Symphony data, live look-up of item status and other issues.

Casalini Libri and @Cult also showed strongly based on their responses, but in a different way. Much of the work related to these vendors has been tied to collaborative research projects where UAL is very active (i.e., LD4P and Share VDE). These products show considerable promise for UAL to continue to iteratively develop completely new discovery solutions in the near future, although to date the work has focused almost exclusively on the move to BIBFRAME from MARC. While a move to one of these systems would not fit with an immediate recommendation to replace existing systems, this level of maturity of the projects may come in the near future and can be evaluated on an ongoing basis by the DOT based on recommendations 1.1 and 1.2. The opportunity for continued development of this type is another argument against vendor lock-in at this time.

Environmental Scan

Librarians with responsibility for discovery and related services at the following institutions:

Institution	Rationale for Selection
Duke University Library	• Duke recently underwent a similar discovery strategy review process (see full documentation here), so we were interested in what they found out in their investigation
University of Wisconsin-Madison	• UW-Madison recently launched their coordinated discovery approach, and are in a similar position to UAL as former Blacklight users
Utrecht	• Utrecht Library is famous for being the "library without a catalogue" and their focus on delivery over discovery (video), so they seemed like an important outlier institution to include in the environmental scan
UBC	• UBC Library is a Canadian peer institution to the UAL, and recently developed Open Collections, an appealing discovery interface for their digital collections

Selected institutions were asked the following questions, in addition to a question about their specific approach to discovery:

1. Could you tell us what your discovery strategy currently is? What are the biggest issues and roadblocks you currently face?

2. If you could change one thing about discovery at your institution, what would it be?

Summaries of conversations with each institution are provided below:

Duke

- Duke was motivated to document their discovery environment and assign technical and functional ownership to each of their discovery-related services as the number of repositories and metadata sources under their control had grown exponentially over the years. Duke acknowledged the challenges of having their content across multiple systems, as well as plans to bring these systems together in the future. Rather than thinking of discovery as one system, Duke understands it instead as a suite of tools and services. The mapping process they undertook helped them to visualize the connections between various metadata silos, and outline the flow of information from source to destination systems. At this stage, Duke's focus is primarily internal, however they have plans to conduct user testing, and are currently addressing pain points

UW-Madison

- UW-Madison decided to develop coordinated discovery after being faced with a potentially time consuming and expensive Blacklight upgrade. UW-Madison forked from the Blacklight project and created a series of apps for discovery based on industry standards from the software engineering community. The success of their approach lies in adopting DevOps practices, having a programmer dedicated to data integration, and validation of designs through U/X methods. Based on user testing, UW-Madison's coordinated approach acknowledges the problems created by homogenizing content in typical webscale discovery approaches, and instead separates content into different categories. Categories are linked to one another through Solr's more like this functionality, and search is sticky between each category view. This approach helps to mitigate the issue of users not knowing which category to find a particular content type in. Separate Solr indexes for each category allows for maximum flexibility given varying metadata standards and user search needs

Utrecht

- In 2012, Utrecht made the conscious decision not to invest further in locally developed discovery tools as statistics showed that users were using tools other than those developed by

the library (e.g. Google Scholar, Web of Science, Scopus) more. Utrecht provides information on their website about the best "search engines" for finding academic content, and integrates this into their information literacy program. They shifted their focus from discovery to delivery over a period of 6 years. Utrecht's holdings are searchable through OCLC Worldcat, and they invest significant effort in ensuring that their article holdings are findable/accessible in Google Scholar, and the major article knowledge bases. Utrecht is also working at the network level to improve interlibrary loan, first in the Netherlands, and then more widely in Europe. Utrecht has received very few complaints from users about discovery, although special collections have proven to be very difficult to locate without a local discovery system

UBC

- UBC Library decided to focus their development efforts on standardizing, consolidating, and surfacing digital collections through a custom-built portal developed in partnership with UBC IT. Developing Open Collections first required standardization of content and metadata housed in multiple silos, and a conscious decision was made not to program around data quality issues. Traditional discovery is currently managed by a committee, and for the time being UBC plans to rely on Summon/Voyager, which they feel works well enough for their users. If more resources were available for traditional discovery, UBC would like to do more in this area

Internal Environment

Stakeholders

Stakeholder	Role	Relationship to Change
UofA students and faculty	The primary users of UAL discovery systems	Primary stakeholders in determining desired functionality of discovery systems so that they can effectively obtain the information resources needed to complete their work
Public Services	Responsible for educating users on efficient and effective use of our discovery systems to meet their academic needs. Responsible for providing feedback on evolving and diverse user needs, confusions, and as well as disconnects with other search systems and best practices	Wants discovery to be consistent and reliable in its search strategy functions, results display, and relevance rankings and intuitive for users. Public service staff need to be comfortable and familiar with the discovery systems in order to adequately teach users and effectively offer public services and instruction across a range of methods.

Collections Strategies staff	Responsible for selecting, acquiring, and assessing UAL collections	Wants to ensure that discovery solutions surface the collections UAL owns/subscribes to, regardless of vendor platform, so that collections are well utilized. Also wants to ensure that restrictions on resource use are conveyed to users
Bibliographic Services staff	Responsible for ensuring appropriate high quality metadata is created that connects users with resources	Want to ensure that discovery solutions can leverage metadata in various formats and standards to enable discovery and access to UAL's rich collections. Also want to ensure discovery solutions allow for efficient and effective workflows for adding/editing/enhancing/deleting metadata in a timely way
Digital Initiatives	Responsible for repository, research data, library publishing, and digital scholarship services	Wants to ensure that digital collections are discoverable by users in UAL's discovery systems and that open access material/locally curated collections are not disadvantaged or hidden from discovery
ITS (developers)	Responsible for recommending technological solutions and designs for the system. Also responsible for the development of the application with proper automated tests. Responsible for designing and developing the application following guidelines/practices to ensure usability, accessibility, supportability and security	Wants to ensure that discovery solutions are based on industry standard practices and tools, to ensure that changes they make to the application yield expected results, and that regular progress can be made on the application without the need for significant rework. This stakeholder group also wants to ensure that the problem of technical debt is accounted for, and avoided as much as possible through proper system design
ITS (system administrators)	Responsible for providing and maintaining infrastructure to support the application. Responsible for working with ITS developers to implement and deploy the application according to a regular release schedule. Responsible for monitoring the	Would like discovery infrastructure to be stable and maintainable, deployments to be predictable, and for the infrastructure/application to be dependable, available and secure

	application, and ensuring performance, availability and security	
Strategic Leadership Team	Accountable for ensuring that mission, values and strategy drive services	Wants discovery to be driven by UAlberta user needs, to be dependable and sustainable, efficient to maintain, and an intuitive and rewarding entry point to the rich resources Libraries makes available to the community
Discovery & Web Services Librarian	Responsible for the design and development of discovery services to support students, faculty and other researchers in accessing the full range of collections and services offered through the Libraries and from beyond	Wants to ensure that discovery is operationally sustainable, performs well, meets user and staff needs, and continuously improves to meet evolving user needs

Staff Resources

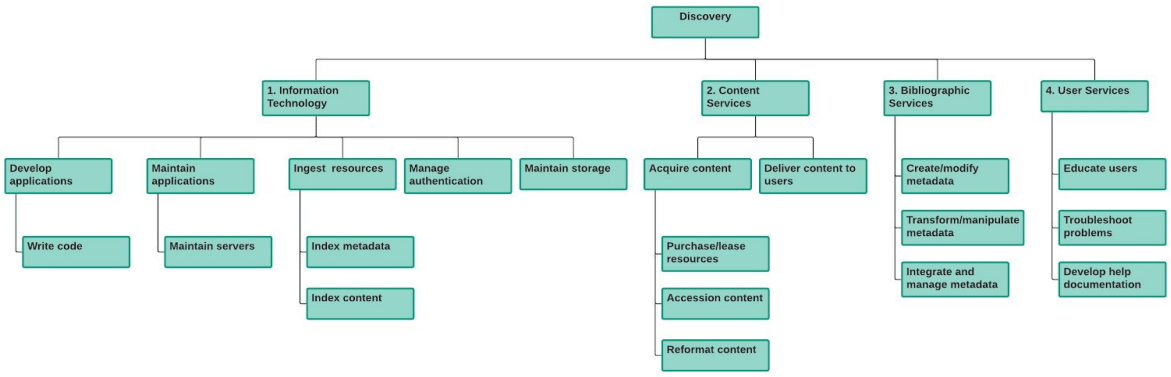
Discovery is currently supported by the following staff resources (estimates):

1. Discovery & Web Services Librarian (1 FTE)
2. Dev Team (currently 0.25 FTE - on a project basis, managed by Digital Initiatives Technology Librarian, 1 FTE)[2]
3. System Administrator (0.15 FTE - as needed for system maintenance, reindexing, monthly deployments)[3]
4. Web Applications Developer (1 FTE - as needed, on a project basis for interface design and usability testing)
5. Various Bibliographic Services staff (FTE difficult to determine)
6. Web Architecture Team (advisory group for discovery - assists as needed with special projects related to discovery)

However, breaking discovery down into its key functional areas through functional decomposition reveals that staff involvement in areas that contribute to the successful operation of discovery can be found dispersed throughout the UAL. As the diagram[4] below reveals, there are many discovery functions

[2] Estimate refers to current resources allocated to discovery development/maintenance by the Dev Team. N.B. from May-September 2019, 2 FTEs were dedicated to discovery development, and on average, about 50% of the Dev Team's time has gone to discovery development since summer 2019.
[3] Estimate based on an average of 2 days/month. N.B. Time required for the initial Blacklight setup was much more significant, and this number increased for significant upgrades/maintenance work
[4] *View full diagram here: https://www.lucidchart.com/invitations/accept/b297e9bc-8f95-4c10-90d6-e621c81decc8

that currently do not have active involvement or input, and discovery functions are dispersed across a number of library units.



Usage Assessment

For the March 2018 sample period, the primary discovery interfaces[5] (UAL and vendor controlled) had the following usage:

1. Blacklight Discovery: 162,679 page views
2. Classic Catalogue (eLibrary): 61,293 page views
3. Databases A-Z list (SpringShare LibGuides): 28,555 views
4. Articles (EBSCO Discovery Search): 55,724 search clicks; 426,956 total requests (Count represents all user activity related to retrieving or viewing, or linking out from a record - see definitions)
5. EBSCOhost: 322,191 record views ; 546,542 record clicks
6. Web of Science Core Collection: 15,136 record views; 33,023 result clicks[6]
7. Scopus: 14,974 record views; 11,261 result clicks
8. ProQuest: 7529 record views ; 12,264 result clicks

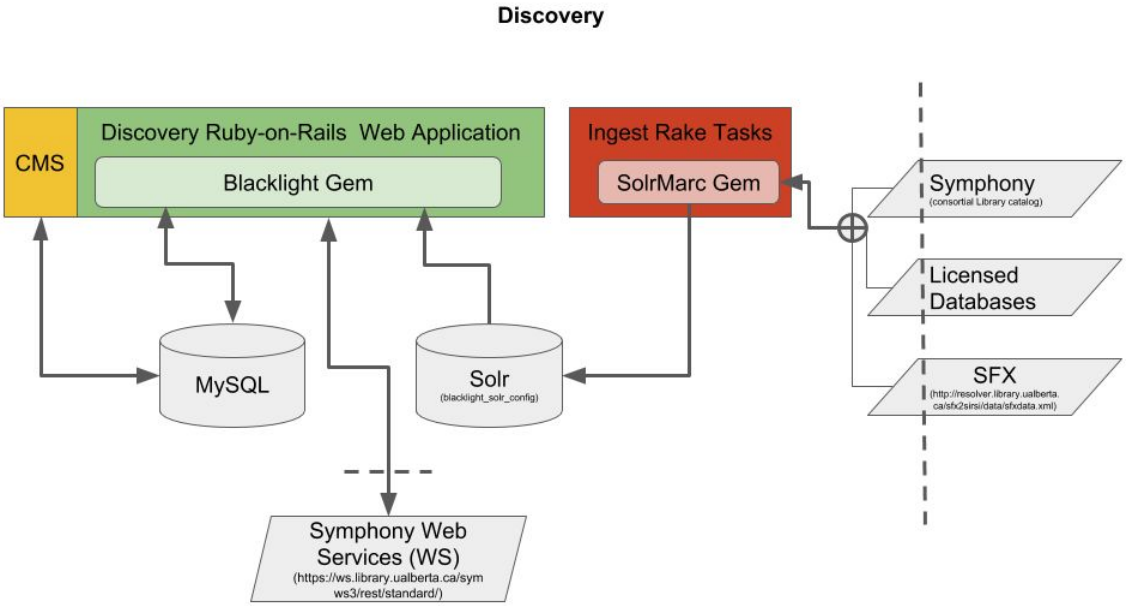
It is important to note that usage in each of these systems has different meanings, so numbers are not easily comparable. It is clear from this usage assessment, however, that the various discovery systems available to UAL users are heavily used.

[5] Google Scholar was on our list of primary discovery interfaces to assess for usage. Given the difficulty of parsing EZProxy stats, however, we were unable to include it in this usage assessment

[6] Result Clicks count all the clicks originating from the result list displayed by a search or browse. This includes links to external resources, as well as records included in the databases on that platform. Record Views cover only views of the detailed metadata (e.g. an abstract view) of records from databases on that platform." https://www.projectcounter.org/quick-guideresult-clicks-record-views/

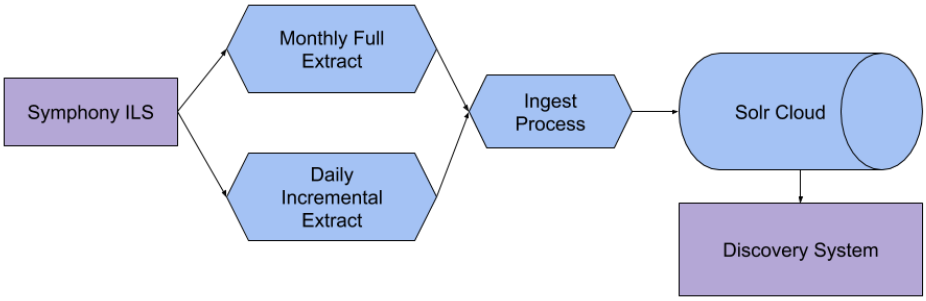
Technical Environment

Discovery Architecture



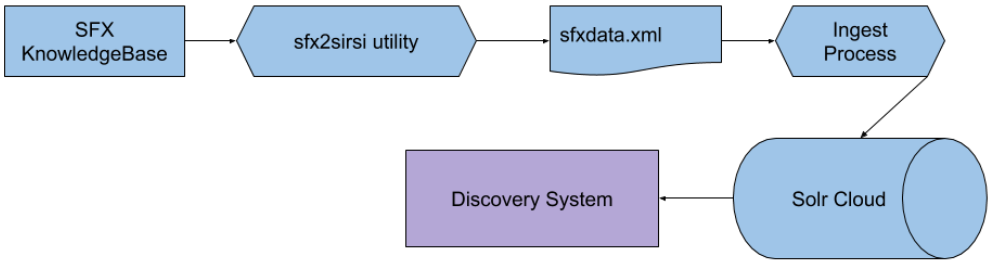
Ingest Processes

Discovery Symphony Ingest Workflow



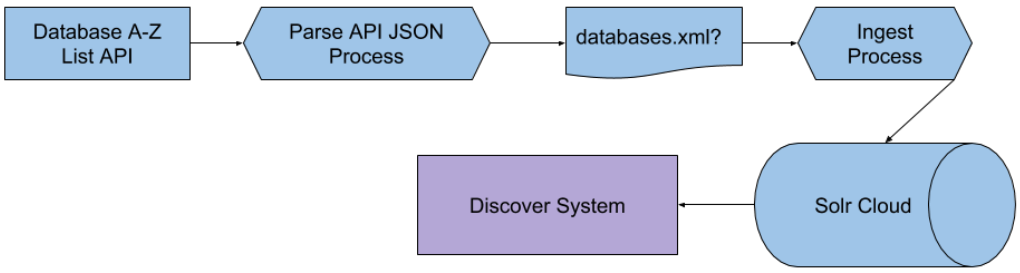
- Both monthly and daily incremental extract are run on the ualapp server.
- monthly full extract script is at /u/sirsi/NonWFcustom/EBSCO/Monthly/ebsco_extract.full.ksh, run every second Sunday of the month
- daily incremental extract script is at: /u/sirsi/NonWFcustom/EBSCO/Incr/ebsco_extract.daily.ksh, runs every day of the week, except for the second Sunday of the month
- Once extract is completed, the export files will be ssh-ed to york, a discovery application server.
- Daily ingests will be triggered by cron jobs
- Monthly ingests are started manually. A full reindex takes 3+ hours to complete.
- Ingests are run through rake tasks: https://github.com/ualbertalib/discovery/blob/master/lib/tasks/ingest.rake
- All ingest related configuration is at: https://github.com/ualbertalib/discovery/blob/master/config/ingest.yml

Discovery SFX Ingest Workflow



- SFX exports the e-journal holdings every night, in an marcxml file format
- Sfx2sirsi (https://github.com/ualbertalib/sfx2sirsi) downloads the marcxml and downloads the summary holdings (date coverage) data, and combine them together to create a complete export of sfx data at: http://resolver.library.ualberta.ca/sfx2sirsi/data/sfxdata.xml
- Cronjob on Discovery app server (york) triggers the rake task to ingest sfx data. Endpoint is configured to point to the file above.
- It removes all existing e-journal data and re-ingest them. It takes 3+ minutes to complete.
- All ingest related configuration is at: https://github.com/ualbertalib/discovery/blob/master/config/ingest.yml

Discovery Database Ingest Workflow



- Cronjob on Discovery app server (york) triggers the rake task to ingest database data.
- The rake task will fetch database list from SpringShare's API in JSON format, and restructure the JSON file into a database.xml file
- It uses the batch ingester to ingest from the xml file: https://github.com/ualbertalib/discovery/blob/master/lib/ingest/batch_ingest.rb
- It removes all existing data and re-ingest them. It takes 1+ minutes to complete.

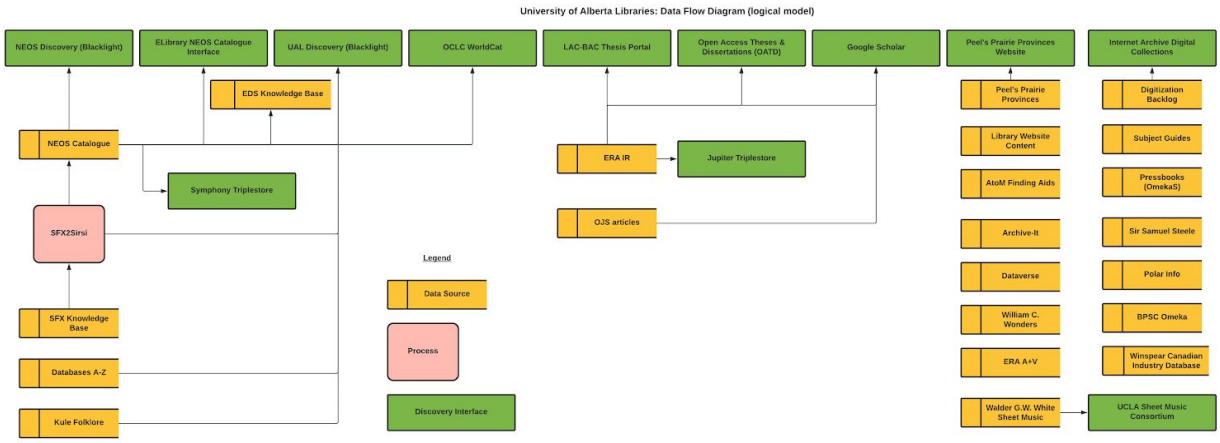
Technical Assessment

Experience with the Blacklight codebase has shown it to be difficult to develop given how it is architected and due to technical debt that has accumulated through the duration of the project. What this means in practice is that bug fixes and the addition of new features that improve the user experience are time consuming and may result in unintended consequences. The codebase has also been heavily customized to support our local requirements, which prevents us from upgrading the platform to later versions of Blacklight without significant rewrite efforts. This offsets the benefit of staying with the Blacklight community as we do not benefit from any new community developed improvements. Designed to be a turnkey discovery solution and to encompass user requirements from the broad user community, Blacklight has many built-in features that do not meet our local needs and add a significant layer of complexity when developing and maintaining the application, as well as potential performance and security concerns. In many ways, the Blacklight codebase has shown to be a barrier to moving discovery service forward in a timely fashion. A symptom of this issue is the long onboarding process for new developers, as they are required to learn Blacklight-specific approaches instead of being able to use industry-standard methods commonly used in the software development community. This problem is further compounded by a lack of separation of concerns between key components of the Blacklight architecture (indexing, ingest, UI), which make software deployments and reindexing more time-consuming than is required to meet operational and user needs, and causes difficulty when diagnoses the root cause of problems. With the ingest pipelines tightly coupled with the application, and the index platform (Solr) inappropriately used as the datastore, it is challenging to

manipulate data mapping and cleaning to improve user access to new collections. All of these factors contribute to a lack of long term sustainability and supportability of the current Blacklight-based discovery application.

Metadata Assessment

Technical challenges with the current Blacklight platform make it difficult to incorporate non-NEOS/UAL MARC based collections (e.g., Kule Folklore) without substantial work either by UAL to create custom mappings, or by partners to make their data conform to UAL/NEOS specifications. This is counter to the desire to provide a lightweight discovery solution for UA collections. In addition, these technical challenges have stalled the processes of incorporating non-MARC based collections into discovery as originally desired and planned for. This has meant that many of UAL's special and unique collections are siloed in a way that hampers effective and efficient discovery and access. As well, the challenges with incorporating non-MARC based collections have contributed to unnecessary duplication of efforts (e.g., metadata created for the same resource twice) and duplication of metadata (e.g., different metadata for same resource in two places). The former takes away resources that could be used for metadata creation elsewhere; the latter puts additional pressure on processes of metadata synchronization and makes update and enhancement more difficult. The move toward linked data that has been happening in UAL and which is increasing in pace and scope also poses challenges in this environment as we will need to accommodate emerging standards such as schema.org, BIBFRAME, JSON-LD, etc. The diagram below[7] demonstrates the current flow of metadata sources in to discovery and highlights the large number of local collections not indexed in discovery:



User Engagement & Staff Adoption

Many library staff and some end users have been hesitant to adopt the use of Blacklight Discovery into their work, continuing to use the Classic Catalogue (SirsiDynix eLibrary) particularly for known-item searching, or the EBSCO Discovery Service (EDS) interface. This has been largely due to a

[7] View full diagram here:
https://www.lucidchart.com/invitations/accept/c3ab5be5-ce72-454e-89d1-2a4b914ce40c

mismatch between some users' desired functionality not being met by Blacklight Discovery and a lack of understanding and confidence in its relevance ranking. An attempt to address these usability challenges in August 2018 through the removal of the Bento UI design was met with a different set of challenges, as Blacklight Discovery until then leaned heavily on the EDS API to present article results. While the EDS interface is currently also available to users as a discovery system on the library homepage ("Search Articles"), its interface is not consistent with UAL's branding, it provides extraneous content (e.g. research starters), and it is limited in how it can be customized. It is not readily transparent which resources are searched via EDS, leading users to assume mistakenly that it searches all content within UAL collections. An extensive Discovery Usability Study, concluded in March, 2018 highlighted the following high-level issues with discovery at that time, a number of which have since been resolved:

1. **Bento Design**: this UI design (since removed in August 2018) confused users, primarily due to the fact that the items were often duplicated across different Bento containers, and that it added an extra UI view in addition to the expected results and records views. The removal of this view caused some additional confusion for users used to seeing article results in their searches
2. **Advanced Search Functionality**: users were confused about why we had two advanced searches (the second was removed in August 2018), and what they could accomplish in an advanced search
3. **Electronic Resources Workflow**: users were confused by the multiple interfaces they were required to view in order to access an item, and did not understand the SFX menu (the SFX menu has since been rebranded)
4. **Record View UI**: a number of consistent usability issues arose due to the information/visual design of record views, which made it difficult for users to locate important information and access features (new record view implemented July 2019)
5. **Facets**: users were able to apply facets, but were often confused by terminologies used, content types included, and consistency between facets used in different views (the Articles view was removed in August, 2018)

End User Focus Groups

As can be imagined, seven hours of semi-structured conversations with seven different UAlberta Library users representing various faculties and stages of academic career resulted in a lot of thoughts, observations, reflections, and suggestions.

Overall, the following themes were identified.

Wide variety of research needs

- Even within our small pool, all participants wanted different types of resources and they all arrived at those resources differently. Participants expressed the need to flip between what is currently available and what is out there beyond our collection. Users have a wide variety of research (e.g. topics, discipline, approaches, languages) and need to be aware of cutting edge research. Some fields require a wide-variety of disparate sources including and well-beyond the most popular works. No participants used only one kind of source.

Desire for simplicity but need for comprehensiveness. (Needs not necessarily self-determined)

- Almost all participants articulated a desire for simplicity but, when probed further, highlighted a need for comprehensiveness in many situations. Navigating these seemingly incompatible concepts throughout their academic careers, users develop more sophisticated search strategies when they discover that they need more than offered by a simple search or one system. Furthermore, several mentioned that their "needs" are often governed by the expectations of others (e.g. instructor, marker, peer-reviewer, journal editor) and sometimes only realized after receiving a low mark or an article/grant rejection because their sources were not considered up to par by their audience.

Desire for more transparency of search tools due to unclear labelling (especially "search the library")

- Participants strongly articulated the need for transparency with regards to the Discovery Service. All were negatively surprised when they learned that the main search box (i.e. "search the library") was **not**, in fact, the most comprehensive search option that searched everything we have access to but only a smaller subset that notably does not include academic articles (except by journal title if that is known). They were similarly disappointed to learn that "search for articles" (i.e. EDS) only indexed a small percentage of our databases. Whether users had learned this previously or during our interview, they all felt this had prevented them from finding the best resources for their projects.
- Users suggested improving transparency through clearer naming of the search boxes or possibly via prompts or hover-over features. To all users, the phrase "Search the Library" implied the most comprehensive search of all resources owned or licensed by the library (even when they have learned it is not). If a comprehensive search is not possible, it was strongly suggested that we change the name of the search box. (note: subsequently, the box has been renamed "start your search" but this likely implies results that give a broad context representing a wide variety of disciplines and genre/formats of materials, so this possibly introduces confusion in a different way).
- Some users indicated the cascade of options in GetIt/SFX from most to least easily accessed wasn't clear and also agreed that an option to "recommend a purchase" might go well here. It was also suggested that a discovery system that ran parallel searches in multiple databases or collections would be a benefit.

Need for variety of search methods

- All participants expressed that their research involves a balance of locating "known" items (citations, specific articles) and "topic" searches (keyword searches to better understand a topic or concept). While all users suggested they mostly used topic searching (accounting for about 75% of their searching), they all related how important "known" searching is for finding other works by an author and tracking down references and citations for many reasons (e.g. check evidence, get more background, more detail on methodological approaches).
- Within topic searching, users identified or agreed that there were times when they needed a broader context encompassing a variety of genre/formats across many disciplines and other times when they needed to search more specifically within a certain discipline or for a certain type of genre/format.

- While not necessarily identifying “metadata” searching per se, some participants indicated that they might search for all works by a given author, or they knew a certain series from a publisher, or that they appreciated subject headings either for their hyperlinks or usefulness in constructing a topic search, or that they went to the stacks and searched for related books around the call number of another work.

Need for learning how to improve search strategies over time

- Participants commented on the need to somehow become more aware of how to find better sources as they progress and scrutiny of their sources by others becomes more important. Some develop more sophisticated search strategies through learning by doing, colleague/supervisor suggestions, by accident or serendipity. Several commented on using Google Scholar (not regular Google however which may be significant) not necessarily because it was better but for familiarity, ease of access, and sometimes more useful descriptions in results displays. They recognise our tools are better but are less sure how to find and use them. All recognised the value of getting out of old habits to develop new skills and awareness of alternatives - some of this could be accomplished through labelling and prompts within the discovery system.

Participant selection and guiding questions

Feedback came from seven academic users representing various levels and disciplines recruited from the “library feedback/user testing group” list. There were 3 PhD (educ, ALES, REES), 1 Master’s (Anthro), 3 faculty (MLCS, Arts/St. Joseph’s, Science); no undergrads were available but the other participants incorporated both their own undergrad experience and their expectations for, and experience with, undergrad classes they teach. Six people were interviewed in person (semi-structured) and one replied by email based on the following questions.

1. What formats or genres of resources are you expected to use for your research / assignments?
2. How is your library/source research for assignments, research, or teaching distributed between finding known items from citations or recommendations and unknown items through topic searching?
3. What processes do you use (whether via our library site or not) to discover and acquire these resources?
4. What does “search the library” mean to you? What would you expect to find in a box with this title?
5. How familiar are you with the various parts of our library site? e.g. the variety of tools and how each works differently from others? Would instructional prompts within your searches be useful?
6. What works well in our system? What are frustrations? What changes to the library site might you suggest to make it more usable?
7. Given that a single search box to search across all resources is not possible, how could a library search page be organized?

UAL Staff Focus Groups

Similar to end user focus groups, several hours of discussions with staff who act as an interface between the discovery systems(s) and users also resulted in too many thoughts, observations, reflections, and suggestions to fully list here.

The following themes seemed to emerge:

Need for library main webpage to make discovery options clear

- Expression of a consistent and strong desire to have a main search box that searches articles, as the removal of this feature has caused confusion and frustration for students and staff. The idea of not having a search box at all (à la Utrecht) was deemed confusing, results in too many clicks, and is likely not feasible given our much larger breadth of collections. Several comments were made about the website already being too cluttered. With the main search box, frustration was also expressed regarding the results ranking algorithm; users indicated the top results are often not what they would consider to be the most immediately relevant to them nor the most current and sorting by date interferes with relevance of results. Of course, this highlights the challenge of relevancy rankings where user expectations are so varied and idiosyncratic. There is also a desire for UA holdings to be displayed first.

Wide variety of user needs, expectations, and approaches

- It was recognized that there are differences in approach used when searching for known items versus starting a literature search with a broad topic in mind, and that having functionality to meet both needs would be advantageous. A simple search box will be particularly useful for students for whom library research might be entirely new (eg. international students). Clear labelling and naming is essential so that users will be directed appropriately to different areas of the website, and visual cues would also be helpful in differentiating types of resources.

Recognition of educational and instructional roles

- It was articulated that UAL staff must continue to serve an instructional role and educate users regarding which resources to use for different purposes, but also that directional prompts could be embedded in discovery to help searchers follow different paths. LibGuides were mentioned as useful, but sometimes overwhelming with too many links. There is wide variability in quality and comprehensiveness, so more consistency is recommended.

Discovery is an ongoing issue

- Staff recognize the importance of our discovery service and web page as the point of contact with users and articulated the need for a dedicated team to oversee the increasingly complex discovery ecosystem.

<u>Participant selection and guiding questions</u>

Four focus groups were held with UAL staff members with relevant responsibilities and experience: 2 for PSAs on service desk/chat, and 2 for librarians engaged with information literacy teaching, research consultation, as well as service desk/chat. Finally, a phone call was held with the chat reference coordinator.

Discovery Phase II Concept Design

Based on current state analysis, DSRT determined that a modification of UW-Madison's coordinated discovery approach would serve the UAL well as it develops the next phase of its discovery systems. This approach provides a high level of flexibility for the UAL, and would allow us to avoid issues of lock-in associated with going with a more monolithic approach. It would also remediate a number of usability and operational issues with the current discovery system, most notably the confusion experienced by users in not knowing which category a particular content type belonged to, and the challenges in returning relevant results when metadata and search relevance configurations are homogenized.

DSRT recommends the following preliminary design for discovery phase II, which would consist of a number of sub-projects as outlined in the objectives section, and would need to be verified through U/X design methods:

- A "coordinated bucket" approach, with the possibility of sub-views for important content types, all connected via central data warehouses. Separate indexes and indexing profiles/views/UI designs for the following buckets would be brought together in an appealing interface:

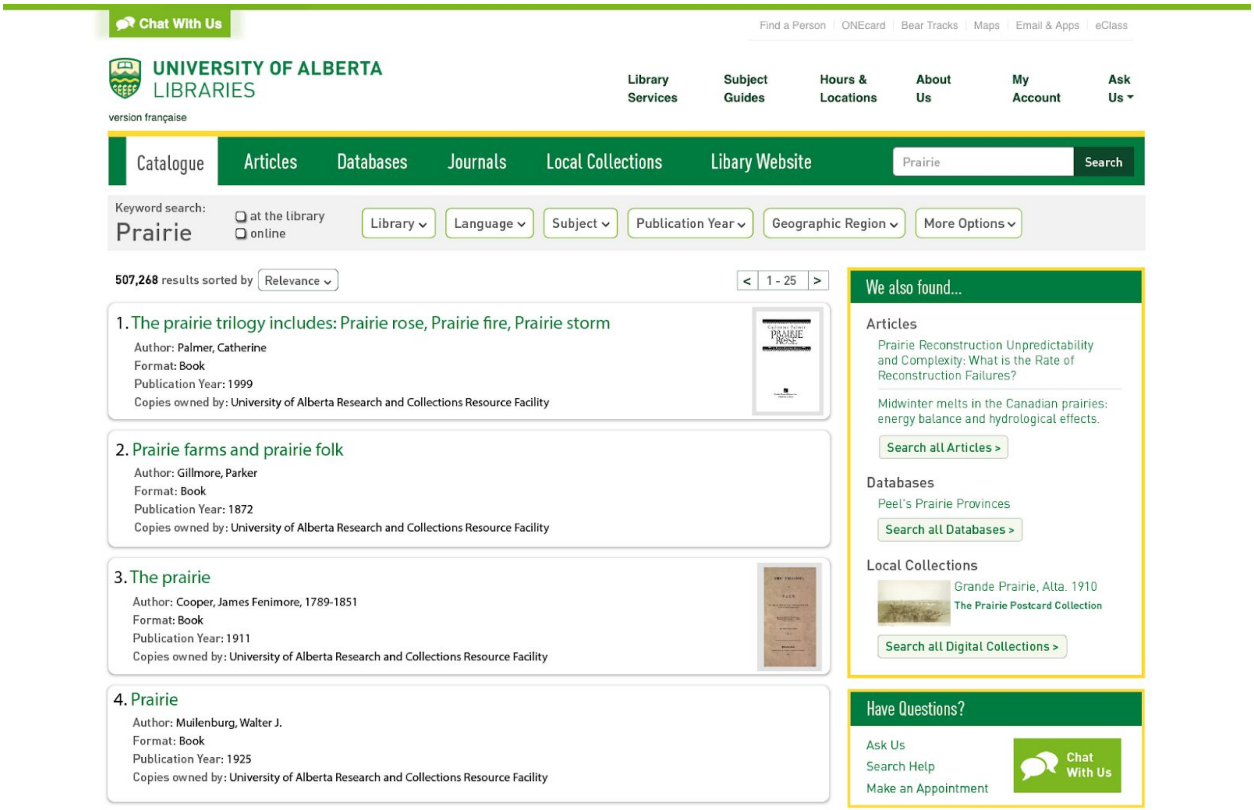
Bucket	**Contains**
Catalogue	• NEOS catalogue, less e-resources with no UA holdings, and content types found in other buckets
Articles	• EDS API results for articles/article-like content
Databases	• Springshare databases A-Z list, including all metadata in subject taxonomy
Journals	• Print and electronic journals from NEOS catalogue/SFX, deduplicated to prefer catalogue record, and ideally merging of print and e records/holdings where possible/appropriate in the discovery view
Local Collections (NEW)	• High impact unique and local UAL collection items not currently indexed in discovery, including digital collections
Library Website (NEW)	• Library website content, including Springshare content (LibGuides)

- A search in a given bucket would return relevant results in other buckets in a "we also found" sidebar. For example, a search in the catalogue bucket would return relevant databases, as users may not be aware that we do not catalogue databases

- As content in each bucket comes with different metadata standards and user requirements, look and feel and application behaviour can differ, as appropriate. For example, facets specific to the databases A-Z list could appear that would allow users to browse for databases using the subject taxonomy developed by the CSU. Similarly, the digital collections bucket could have a record design that incorporates an image viewer
- The visual design would be based on UAL's new brand guidelines, with a strong focus on good web accessibility practices, and simplicity
- Contextual information about what users are searching for in a given bucket would be readily available, in addition to prominent placement of AskUs!
- Persistent links to services (e.g. ILL, Book an Appointment, etc.) would appear on the search results pages
- Robust and varied browse functionalities would be developed for the Catalogue bucket, and other buckets as needed, to provide users with different information seeking methods making full use of metadata linkages (e.g. virtual shelf browsing since physical shelf browsing is less possible with offsite storage).
- Special effort would be put into highlighting local/special collections though UI design features, and search relevance ranking manipulation (e.g. sponsored results)
- The inclusion of more access points than just text (e.g. most use of images, etc.)
- Improved search relevance through regular search log analysis, and potentially the use of machine learning models such as Solr's learning to rank (LTR) plugin
- Introduction of standard search interface features such as user controlled options for spelling corrections, autocomplete, "did you mean" , hit highlighting, etc.
- Search alert functionality to notify users of recently added material and new acquisitions in all formats.

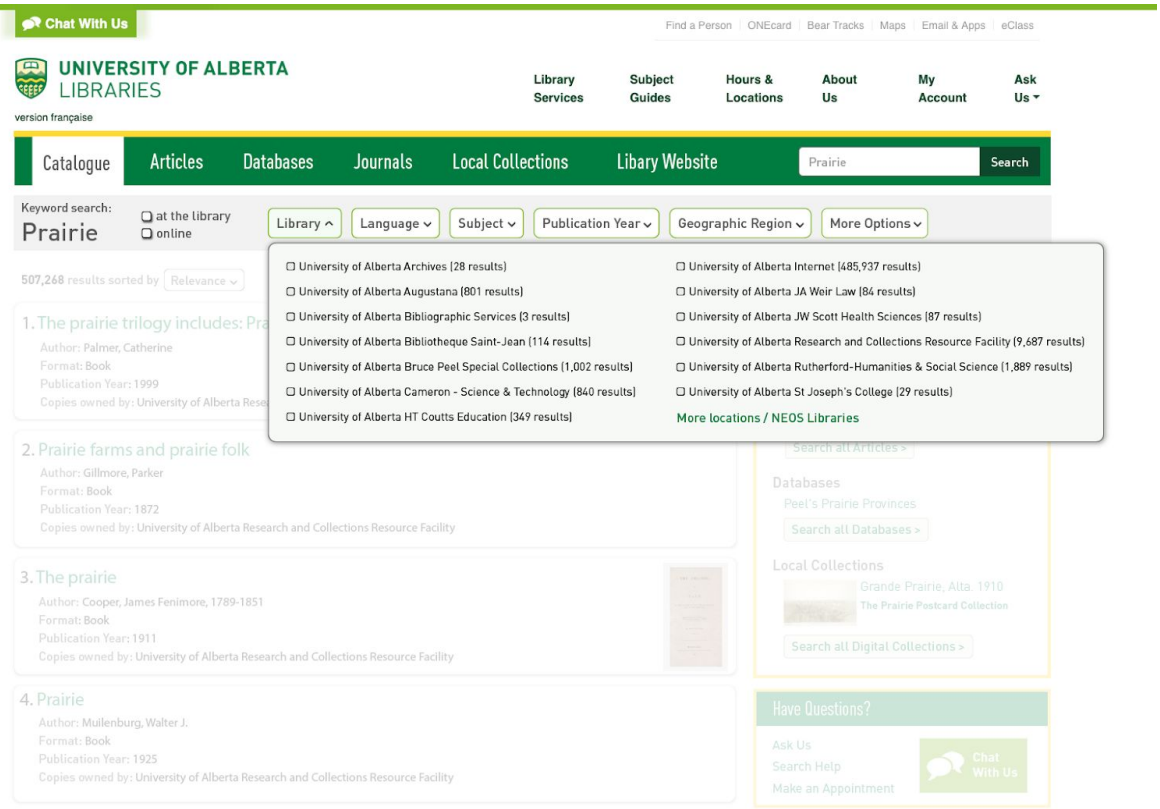
*For more details on discovery phase II functionality, see Appendix 1: Discovery Phase II High Level Requirements

UI Mockups



Mockup 1: Search results page showing new visual design focused on new branding design, web accessibility, and coordinated bucket approach

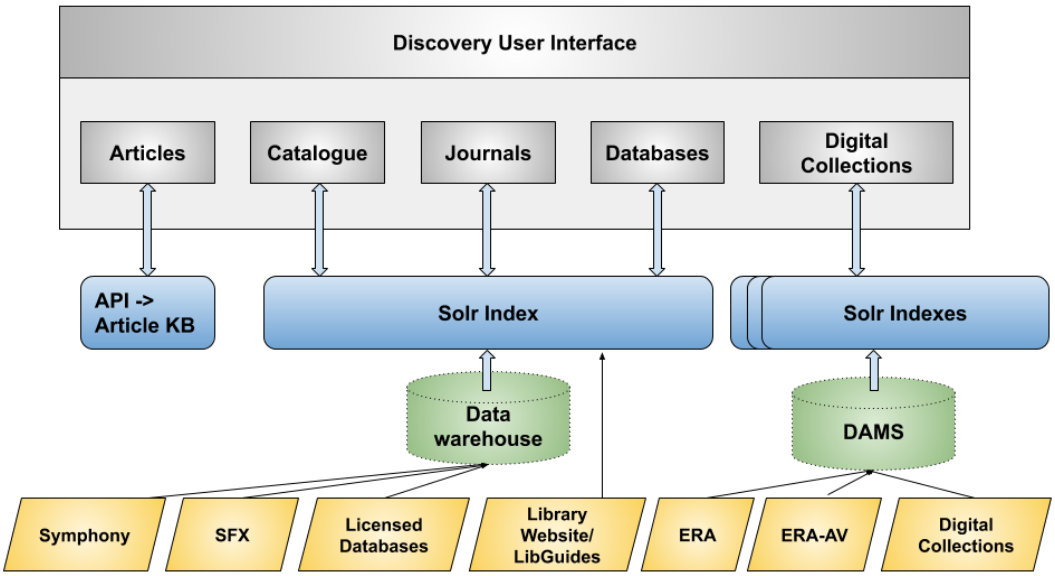
Mockup 2: Main search box with limiters for different buckets, and pre-search filters for different fields.



Mockup 3: Facet selection

Technical Approach

The preliminary technical architecture for the "coordinated bucket" approach can be designed as indicated in the diagram below:



- Article search will be based on asynchronous API requests sent to a selected vendor article knowledge base (e.g. EDS API). The asynchronous design will ensure the search result page will be more responsive to user queries. The impact of performance and availability of the API service on the page response time will be minimized.
- A central feature of the architecture is a data warehouse and comparable data store structure such as our Digital Asset Management System data store. Separate pipelines will be refactored or built to ingest data from various source data stores, including Symphony, SFX, Springshare Database A-Z list, LibGuides, Library Website, local research collections and other source data stores. A similar approach will be taken for our digital assets, including ERA, ERA-AV, Dataverse, digitized collections, etc.
- The data warehouse is a system that pulls together data from many different sources within our metadata ecosystem. It will provide a consolidated data flow to the Discovery interface. The benefit of a data warehouse is the ability to clean, shape, consolidate, and manipulate the complex metadata from various sources before they are indexed and presented to the end-users. It enables experts in cataloguing and metadata to have ownership over the quality of the data in the Discovery system. It is anticipated that the Data Manipulation Specialist (see Recommendation #3) would play a critical role in the design, operation, and maintenance of the data warehouse
- Data from the central data warehouse will be ingested into one or multiple Solr indexes, which will drive the search results presented on the user interface in various "buckets." The same search requests will be sent to multiple Solr indexes simultaneously to power the "We Also

Found…" functionality on the search result page. This will enable a seamless user experience when switching between different "buckets"

- The design of the data warehouse and Solr index(es) including data model and structure, the technology selection for various components within this architecture, as well as the integration with the Digital Assets Management datastore will need to be evaluated and investigated in further details in separate sub-projects
- This technical design maximizes the flexibility of our Discovery system while providing a consistent user experience. It removes the dependency of the discovery system on any specific platform or technology, as it will be composed of smaller, more modular components that can be more easily replaced. It enables us to improve the quality of the data for the Discovery system before the data are indexed for search. And It allows us to be more agile in terms of data modelling, indexing rules, and relevance ranking.

Operational Approach

A successful discovery solution requires the cooperation and input of staff from units across the library system, as this function impacts the day-to-day work of nearly all staff in the UAL in some form. Our current organizational structure and functional segmentation of key elements of discovery mean that there is a lack of end-to-end ownership and understanding of the processes that make up discovery, and no established mechanism for staff to meaningfully contribute to improving discovery. This has resulted in challenging problems to diagnose, insufficient communication, and a lack of collective ownership/endorsement of discovery and its promotion to users. A significant challenge that has underpinned this is a lack of consistent dedicated staff resources to maintain and develop discovery. Discovery has in many ways been operated on a project basis, with resources borrowed from other teams as needed. DSRT determined that the establishment of a Discovery Operations Team (DOT) would help mitigate these issues, and help ensure that discovery has the regular resources and input required. Below is a draft terms of reference for this new operational team for discovery.

*Please note that the terms of reference below were developed before the announcement of the emerging functional team-based model in development at the UAL, so are subject to significant revision. It will be important to develop DOT in a way that allows it to succeed in the new model:

Discovery Operations Team (DOT) Draft Terms of Reference

Role

To manage the ongoing maintenance, development, improvement, and assessment of library discovery services, broadly defined

Specific Responsibilities

- Gathering and developing user requirements, conducting usability testing

- Soliciting and responding to user and staff feedback
- Developing use cases and prototypes, wireframes, design concepts
- Software testing/quality assurance
- Liaising with the UAL staff (dev, and sys admin, bib services team members) assigned to discovery, where areas of concern overlap (e.g. ILS team)
- Continual assessment of open source and vendor-based discovery products
- Documenting issues and improvements to discovery in GitHub
- Sub-groups may be tasked with specific responsibilities (e.g. assessing information provided by vendors, usability testing projects, etc.)

Membership

- Discovery & Web Services Librarian (Chair, ITS)
- Digital Initiatives Applications Librarian (ITS)
- Data Manipulation Specialist
- 2 representatives from Bib Services (metadata & cataloguing)
- 2 representatives from Collection Strategies Unit
- 2 representatives from Public Services

Commitment

- During active development/product assessment - weekly
- Regular meetings every 2 weeks
- Additional time commitments for special projects as they arise
- May include responding to user/staff feedback, answering help desk tickets, etc.

Anticipated Duration

- As DOT is an operational team, there is no anticipated end date. Terms of reference will be reviewed on an annual basis

Method of Operation

- DOT is an operational team, so members are expected to contribute to the maintenance and development of discovery on a regular basis, as a significant aspect of their duties
- DOT has the authority to make changes to discovery, but will consult with relevant committees and individuals in the UAL where appropriate

Reporting Structure

- The Discovery Operations Team reports to SLT

Goals

The Discovery Phase II Concept Design presented in this report will allow UAL to achieve the following high-level goals:

- **Goal 1**: A discovery solution that is designed based on early and ongoing user-centered design methods, consistency with other commonly used discovery systems, requirements elicitation and validation, and formalized testing/quality assurance procedures, to ensure users are being well served and discovery applications are stable
- **Goal 2**: A discovery solution that allows UAL to more easily integrate metadata housed in different information silos, and to present it attractively to users, regardless of format, in order to highlight unique to UAL resources
- **Goal 3**: A discovery solution that makes it easier for developers and system administrators to make changes to the system infrastructure and application on a regular basis, in order to respond quickly to bugs and develop valuable features for users
- **Goal 4**: A team-based approach with sufficient staff resources to support the ongoing development, maintenance, and continuous improvement of discovery that streamlines the management and promotion of this service, and increases user engagement

Objectives

The UAL can achieve the 4 high-level goals by completing the following objectives:

Objective	Description	Details
1A	Increase user adoption of and satisfaction with discovery	User survey reveals high satisfaction rate (3.5/5 and higher), usage of Classic Catalogue and EDS declines relative to discovery
1A1	Improve search functionality	Relevance/precision testing aligns with user and staff expectations, and we are taking advantage of modern indexing technology features
1B	Initiate user-centred design process for the next phase of discovery	Meet with end users and UAL staff to get feedback on preliminary design concept. This will include the creation of mockups, wireframes, usability tests, etc.
1C	Create list of functional/non-functional discovery system requirements	Complete a list of requirements that has been vetted by UAL stakeholders. Based on requirements, create solution design (architecture design, solution selection, etc.)
1D	Develop discovery test/QA plan	Work with UAL technical stakeholders to articulate formal discovery system testing plan (including list of automated tests to be created, UAT scripts, etc.)
2A	Create UAL discovery metadata requirements	Work with Bib Services staff to create metadata requirements (index/facet/display) for discovery for various content sources

		(Symphony, digital collections, etc.). Determine design and architecture requirements for metadata solution for Discovery Phase II
2B	Create UAL discovery metadata migration/cleanup plan	Work with Bib Services staff to create metadata migration/cleanup plan, which articulates required work and timelines necessary to ingest metadata in discovery phase 2
2C	Ingest high priority digital collections metadata into discovery phase 2	Work with Bib Services and Library Publishing to prioritize digital collections metadata to include in discovery phase 2
2D	Recruit/backfill Data Manipulation Specialist	Recruit or redeploy a Data Manipulation Specialist who will ultimately be responsible for the discovery ingest process, associated systems, data cleaning and transformation scripts, etc.
3A	Refactor discovery codebase	Work with dev team to determine areas of current Blacklight code that should be removed and/or cleaned up and remediate technical debt necessary to implement discovery phase 2 requirements. Continue the work done to maintain the aspects of Blacklight code that work for UAL, and simplify those areas that do not. Following the approach taken at UW-Madison, for the time being, stay with the Blacklight architectural approach while other solutions are investigated
3B	Implement deployment process improvements	Work with dev and sys admin teams to identify smoother deployment processes, the ability to better support same day hotfixes, and implement necessary changes in applications and infrastructure to support this objective
3C	Develop/implement improved ingest processes	Implement ingest process that is not coupled with the discovery application code, does not require deployments, and allows for expedient loading of new collections
4A	Form operational group for discovery	Drawing on current and future staff resources dedicated to discovery, form an operational group that meets on a regular (at least bi-weekly) basis to identify issues, determine improvements, respond to user feedback regarding discovery, and approve changes
4B	Assess vendor-provided responses	By the end of 2019, assess vendor responses to discovery requirements (see Recommendation 1.1)
4C	Create discovery system roadmap	Working with end users and consulting with other UAL staff, develop a discovery system roadmap that outlines maintenance and enhancement plans
4D	Increase staff awareness, training,	Develop staff training materials, and offer information sessions on

	and education around the discovery service	discovery

Appendices

Appendix 1: Discovery Phase II High Level Requirements

Appendix 2: DSRT Vendor Responses Summary