Discovery and Access
Recommendation and Report

CUL 2011
Contents

Executive Summary ........................................................................................................................................ 3

About the Appendices: .................................................................................................................................. 8

Appendix A: CUL’s Current Discovery & Access Environment & Defining the Ideal (Group 2) .......... 9
  CHARGE: ...................................................................................................................................................... 9
  Information Gathering ................................................................................................................................ 9
  The Ideal System ....................................................................................................................................... 10
  Gaps and Shortcomings ............................................................................................................................ 11
  Appendix A.1: Specific Recommendations from Focus Groups ............................................................. 15
  Appendix A.2: Dimensions of Discovery and Access ............................................................................... 17
  Appendix A 3: ............................................................................................................................................ 18

Appendix B: Information Architecture (Group 3) ..................................................................................... 22
  Architectural options ................................................................................................................................ 22
  Sources to aggregate .................................................................................................................................. 23
  Use of local data to improve discovery ..................................................................................................... 25
  Aggregation technologies .......................................................................................................................... 25
  Registry of digital collections .................................................................................................................... 26
  VIVO for aggregation ................................................................................................................................. 27
  Population and maintenance ...................................................................................................................... 30
  Exposing data in the aggregation layer .................................................................................................... 31
  Recommendations ...................................................................................................................................... 31

Appendix C: Examining Internal Processes at CUL (Group 4) ................................................................. 32
  CHARGE: ...................................................................................................................................................... 32
  Introduction ................................................................................................................................................ 32
  Landscape of groups ................................................................................................................................. 33
  Case Studies ............................................................................................................................................... 34
Executive Summary

Summary Recommendations

1. Implement a discovery and access solution that integrates a vendor-provided platform for licensed content and selected local resources with an integration layer to manage other local information resources and locally created context (e.g., metadata, virtual collections, relevant subject librarians, annotations) about all Library resources.
2. Implement Summon from Serial Solutions as the discovery system for licensed content and selected local content.
3. Develop an integration layer for local content and context based on RDF data using the VIVO platform. This layer will also serve as a new “digital collections registry”.
4. Create an integrated user experience that presents results from our licensed materials, relevant catalogs, and the new integration layer in a single coherent display.
5. Seek to simplify the creation and maintenance of the integration layer by reducing the number of different silos and separate collections we maintain.
6. Develop an implementation plan that reflects and funds ongoing efforts throughout the entire life cycle of the proposed environment – not just the start-up costs.
7. Address how to track issues in the operation of the discovery and access system, ideally with an enterprise-wide ticket tracking system.
8. Evaluate integrating the BX Recommender scholarly recommendation service with Summon, and acquire BX recommender if acceptable integration can be achieved.

Proposed Implementation

- Create three implementation teams drawing on members of the current Discovery and Access team with possible additional members: 1) Summon implementation team; 2) Integration layer implementation team; 3) User experience team.
- Complete purchase of Summon by July 15, 2011.
- Implement a beta version of Summon that searches licensed article content with no UI customization or local content for fall 2011.
- Expand Summon implementation to include appropriate local resources (e.g., Voyager catalog, eCommons, CULLR) for winter 2012, including a specific focus on meeting the needs of our new virtual libraries. Create a customized and integrated user interface for Summon search.
- Beginning now through the fall design the integration layer and how it will work together with Summon to deliver a coherent user discovery experience.
- Based on the initial design work, by January 2012 propose a specific implementation plan with projected costs for the integration layer for review by LibExec.

Costs

- Assuming a three year commitment, the first year cost for Summon would be $82K, with the second and third year costing $71K each.
• Initial implementation of an out-of-the-box instance of Summon for the fall beta should be straightforward, involving no more than a couple of person-weeks of effort (<$5K).
• Expanding the Summon implementation to include local resources will require significant effort by IT and other Library staff, but will be dependent on the resources included.
• Implementation of the integration layer will also involve significant staff time, which should be estimated during the initial design phase.
• Since these costs are over $25K, a business plan will need to be submitted to the Academic Computing subcommittee of the ITGC, but approval can be assumed.

Overview

The current CUL discovery system has major shortcomings that impede or complicate research. The system lacks a single search box to allow a user to search across all the component systems and yield differentiated source results. Many relevant available resources are difficult or impossible to find, and the overall search functionality of WorldCat Local is sorely lacking. Other problems include, but are not limited to: 1) confusing search results due to widely varying coverage and content from providers; 2) discrepancies among the results returned by aggregators, and 3) search functionalities that differ across the component systems of the discovery system. These add up to a discovery system that frustrates many of our user communities, most notably faculty (some of whom have been quite vocal about their displeasure).

The Discovery and Access Team has been charged with developing recommendations for processes and systems to dramatically improve discovery of and access to digital and physical information resources relevant to the scholarly and creative needs of Cornell faculty, staff, and students. The proposed solution must accomplish the following:

• Break down the current information discovery silos, which require patrons to use a wide variety of different systems and resources to discover potentially relevant materials (e.g., Luna, Voyager catalog, MOA, Hearth);
• Ensure that relevant Cornell and other resources are easily discoverable through commodity search engines, a frequent starting point for our users;
• Ensure that ongoing maintenance for the site be resourced and accounted for within the implementation plan.
• Separate discovery from the underlying technical details of catalogs and repositories of information resources, so that those can easily be changed or added to without adversely affecting discovery;
• Ensure that discovery solutions connect users with the right librarians to support their research and learning, as well as the right information;
• Include discovery of resources that Cornell licenses or supports access to (e.g., BorrowDirect or ILL), as well as resources that we hold; and
• Leverage the selection, description, and categorization work done by librarians (e.g., in LibGuides), faculty (e.g., in course reserves), and users (e.g., anonymized patterns of usage) to provide context for information resources and improve their discoverability.

An ideal system would return relevant results as quickly as Google, be as easy to browse as shopping sites like B&H Photo (http://www.bhphotovideo.com/) and have filtering and customizing features like Amazon. It would be able to access the full range of resources that we
provide to our user community, accommodate different types of discovery from browsing to known item searching, and have help readily available whether the researcher is using a cell phone or a laptop.

In conducting our research, four different dimensions related to discovery and access were considered: 1) the various types of resources (e.g., full text, images, indexes & abstracts, etc.); 2) different user types (e.g., faculty, beginning undergraduate, alumni, etc.); 3) facets of the research process (e.g., reference sources, background reading, primary resources, citation verification, etc.) and help availability, (e.g., consultation with a librarian, subject guides, etc.); and 4) the range of devices (e.g., laptop, mobile, iPad, etc.) that a person might be using. Our users should be able navigate the full range of resources and connect to relevant articles, images, and data on any device plus be able to find background resources (e.g., subject encyclopedia content) and personal help at any stage in the process.

After researching available tools and software, we concluded that no single vendor product provides a viable solution to connect to and index content in CUL's various local collections and value-added information alongside our licensed resources. Furthermore, it will not be possible to build a system where we locally aggregate metadata and the full content for all of our licensed resources (both because of permissions and the impracticality of dealing with very many sources). We must thus develop a solution that integrates discovery and access of content aggregated locally with licensed resources delivered through a vendor service.

We propose creating an integration layer to support effective aggregation, indexing, and use of locally held content and to support the creation of user-focused views. This layer will bring together diverse information sources including repository content, digitized holdings, multimedia collections, and “value added” components such as LibGuides, Course Reserves, Voyager statistics, library web pages, etc. Some of this information will be used to influence rankings and to annotate results rather than simply being indexed as additional items. We recommend that this integration layer be based on RDF data using the VIVO platform. Maintaining support for aggregation from multiple sources requires a registry of collections to aggregate. The current “digital collections registry” is inadequate for this purpose and is based on obsolete technology. We again recommend using the VIVO platform. VIVO can accommodate the migration of existing data and functionality while providing the flexibility to extend the data model to meet the requirements of the new integration layer.

The diversity of our local resources (including, for instance, all our metadata in Voyager, annotations of external resources by CUL staff, locally digitized text and images, and combined local and external resources such as HathiTrust) and the lack of standard data formats and interfaces will require a significant effort both to aggregate them and to expose them to commodity search engines. Work must be prioritized and phased. Going forward we should, where possible, reduce the number of different silos and separate collections we maintain. When considering new technologies for collections, we should ensure that they provide APIs for metadata and full-content harvesting on an incremental basis to support aggregation and indexing.

Selection of a vendor platform to provide discovery of licensed content, alongside our aggregation of local information, should be based on the following criteria: 1) coverage of our licensed content, 2) ability to integrate with searches of locally aggregated content, 3) user interface functionality, and 4) ability to integrate local information into searches to influence ranking and presentation (e.g., annotation). Our survey of vendor products suggests a shortlist of Summon, EBSCO Discovery, and Primo as candidates. After a thorough review of these options, we recommend that CUL license and implement Summon from Serials Solutions. Of the
Discovery Services investigated, Summon offers broad coverage of our licensed e-journal content (estimated at approximately 91%), the ability to integrate with locally aggregated content via APIs or content harvesting, and an up-to-date and evolving user interface. A coverage analysis provided by Serials Solutions showed that of approximately 33,000 e-journals Cornell manages in the Serials Solutions knowledgebase, less than 3,000 are not represented in Summon currently. The API offered by Summon has been used successfully at several academic institutions to provide a “mash-up” including results from Library Websites, local catalogs, and other content. Summon can be implemented relatively quickly, “out-of-the-box,” as a robust article discovery system, then expanded with local collections over time. As CUL develops an integrated discovery layer, Summon will continue to play an important role within the bigger picture.

The ideal discovery system would present results from both the vendor-aggregated content and locally-aggregated content together, merged into one ranked and faceted result set. Some vendor platforms offer limited abilities to import local records, but these currently fall short of what would be needed to include all of CUL's locally aggregated content, although we might experiment with such facilities. The logical alternatives of either providing a locally-built search engine which includes vendor-aggregated content or federating vendor and local search results into a combined result set are not possible because of permissions and technical issues.

We recommend adopting a “mashup” solution as the best way to present our users with one view of results that span diverse local and licensed collections. An example of this approach is found at North Carolina State University (NCSU, http://www.lib.ncsu.edu/). The site uses Summon as a metasearch tool in combination with the in-house-developed QuickSearch tools which presents federated results from Summon, the local catalog, subject guides, local repositories, etc. and presents them as a single coherent display. Features such as autofill and “did you mean” help the user use the single search box easily. Separate display areas for different sources present relevant information, including the location and call number in a brief record for catalog entries. (Better integration of subject guides, reference material, and greater visibility of user help would enhance the system but overall the site is impressive.)

The CUL discovery system should build and improve upon experience from other sites using the “mashup” approach. The integration layer we propose will allow a more sophisticated combination of data, linking, and annotation of results. We aim also to use local information to influence presentation of results within the vendor portion of the results. Finally, the use of a local integration layer allows design flexibility in the choice of what is combined or presented separately. NCSU uses six result panes which is probably around the upper limit for good usability. In development of the CUL discovery system we recommend usability evaluation of different presentation options using different numbers of result panes, item-type combinations, and result formatting.

In order for our charge to be successful both now and going forward, we must address certain cultural issues within CUL. CUL's current organizational culture minimizes the commitments and costs of ongoing enhancement and maintenance and, for the most part, provisions only for development and other up-front project costs. We also need clearer roles and responsibilities for staff working on projects. To mitigate these issues, CUL should encourage and resource grassroots solutions and build a culture of project management based on SMART goals. Maintenance for ongoing services and integration with CUL's existing infrastructure needs to be properly resourced. A successful deployment of a new discovery and access environment will require an iterative effort, whereby we make changes and request feedback from key stakeholders, such as
faculty, students and staff. Hence the implementation plan needs to reflect and fund ongoing efforts throughout the entire life cycle of the proposed environment (not just the start-up costs).

Finally, CUL should integrate an enterprise-wide ticket tracking system into discovery and access operations. A brief survey of the organization's operations reveals an array of different methods used across CUL for ingesting and tracking support for discovery and access services. Consequently, each project implementation group selects its own preferred tool for ongoing support, which means we have an array of different tools in production across CUL. This is obviously an inefficient way of doing business. One strategy to improve the situation would be to mandate use of one preferred system to track operation support and enhancement tickets for all new projects. Once this base system is established, new project teams wouldn't be burdened with figuring out which system to use. Ideally, the same ticket system would be used in the implementation phase of the service life cycle, to enable more seamless transition from project to service. An enterprise wide ticket system that transcends all of our operations would significantly increase accountability and efficiency across the institution.
In order to effectively gather data and make a recommendation for our Discovery & Access environment within a reasonable time-frame, it was necessary to parse out the taskforce into separate teams. Four groups were created, each with an associated charge: Group 1 researched existing software applications and systems; Group 2 examined the current CUL environment and defined the ideal; Group 3 explored the technical challenges to creating an integrated framework; and Group 4 examined the internal mechanisms for developing, updating, and maintaining CUL software and systems related to Discovery & Access. In synthesizing the final report, we have reordered the group reports, starting with an examination of the existing system compared to the ideal, then moved to the technical challenges of creating the necessary infrastructure, and ended with information on the organization challenges of implementing such a system. Research from Group 1 is Appendix D. When necessary we have given additional context to the group reports, but have left in some of the verbiage related to the separate groups.

About the Appendices:

In order to effectively gather data and make a recommendation for our Discovery & Access environment within a reasonable time-frame, it was necessary to parse out the taskforce into separate teams. Four groups were created, each with an associated charge: Group 1 researched existing software applications and systems; Group 2 examined the current CUL environment and defined the ideal; Group 3 explored the technical challenges to creating an integrated framework; and Group 4 examined the internal mechanisms for developing, updating, and maintaining CUL software and systems related to Discovery & Access. In synthesizing the final report, we have reordered the group reports, starting with an examination of the existing system compared to the ideal, then moved to the technical challenges of creating the necessary infrastructure, and ended with information on the organization challenges of implementing such a system. Research from Group 1 is Appendix D. When necessary we have given additional context to the group reports, but have left in some of the verbiage related to the separate groups.
Appendix A: CUL’s Current Discovery & Access Environment & Defining the Ideal (Group 2)

CHARGE:
Identify current and ideal components and functions for end-user discovery and access, building on and refining the earlier work on personas. Create a list of functions needed by our patrons - prioritize by importance to our users (irrespective of technical issues.) Work to characterize the current challenges of both discovery and access by describing how the current CUL discovery and access functionality compares to the ideal.

Information Gathering
Group 2 focused their information gathering in two areas: 1) forums with faculty, staff, and students; 2) background reading and reviews of recent system implementations at other libraries. Because of the short timeframe and the winter intersession, we did not have an opportunity for extensive outreach efforts. The three sessions that we held with staff, faculty, and students, however, were valuable sources of information—both in terms of the frustrations that users are encountering with our current systems and the features that students, staff, and faculty would like to see in a next generation system.

Feedback from stakeholders.
Group 2 conducted a series of forums with staff, students and faculty. In the first session on Dec. 22nd, we partnered with the Reference and Outreach Committee and the Instruction Committee to elicit library staff opinions. A second session was held on February 3rd with thirteen members of the Faculty Library Advisory Board. And, on February 7th, we held a third forum with 14 members of the Student Library Advisory Council.

Staff Forum: We used the University of Rochester’s ethnographic process and conducted three separate exercises. These included: 1) reviewing key components on search and results screens; 2) identifying key features/functionalities from a non-library site; and 3) designing the ideal discovery and access site.

Process: In the first exercise, staff marked-up print-outs of a search screen and a results screen from WorldCat Local. We asked participants to think of an ideal research environment (regardless of what technologies are realistically available) and to cross out pieces or functions that they felt were not needed, change any wording, and add new functions that they felt were needed. In the second exercise, we had staff think of an online site that they liked and used regularly— not necessarily a research site-- and list the features that made it easy to use. Finally, we provided art supplies and asked staff members to design (and draw) the ideal discovery and access environment. We extracted the themes and functionalities from the results and these are compiled in Appendix A.1.

Faculty Forum: On February 3rd, we held a session with thirteen members of the Faculty Library Advisory Board. Because time was limited at the Board meeting, we presented
the faculty with five functions that had been identified in the library staff focus groups and asked the faculty to rank the functions as high, medium, or low priority. The faculty listed the desire for an "everything" or "search all" box with differentiated search results as their highest priority.

**Student Forum:** The third session was held with the Student Library Advisory Council on February 7th. We conducted two exercises with the students: the WorldCat Local Exercise and the design exercise that we had conducted with staff. The students were very enthusiastic in their responses and offered many excellent suggestions.

**Research studies: Literature and reports**

We looked at the ongoing and previous CUL projects that have related areas of interest. Those include the 2009 Discovery and Access cycle, the work of the CUL Virtual Libraries Group, the personas created by the Library in 2007 (https://confluence.cornell.edu/download/attachments/117767326/cul_personas_final2.pdf), and the transcripts from the faculty and graduate student interviews done for the University of Rochester XC project. We reviewed relevant research studies as they addressed criteria patrons use to determine whether to use a journal article. [1] Several libraries, including the University of Washington (see “Lessons Learned: How College Students Seek Information in the Digital Age”); http://projectinfolit.org/pdfs/PIL_Fall2009_Year1Report_12_2009.pdf and the University of Minnesota (see “Discoverability”; http://conservancy.umn.edu/bitstream/99734/3/DiscoverabilityPhase2ReportFull.pdf) recently issued recent reports related to access and discoverability. And, finally, we engaged in a continual review of other library sites and search engines for additional desirable functionalities.

**The Ideal System**

An ideal system would return relevant results as quickly as Google, be as easy to browse as shopping sites like Pajamagram and B&H Photo, Video & Pro Audio, and have filtering and customizing features like Amazon. It would be able to access the full panoply of resources that we have provided to our user community, accommodate different types of discovery--from browsing to known item searching; and have help readily available whether the researcher is using a cell phone or a laptop.

Our group created a list [Appendix A.2] that outlines four different dimensions related to discovery and access. In this grid, we describe the various types of resources (e.g. full text, images, indexes & abstracts, etc.); different user types (e.g. faculty, beginning undergraduate, alumni, etc.); facets of the research process (e.g. reference sources, background reading; primary resources, citation verification, etc.); help availability, (e.g.

---

1 Barry and Schamber 1998, Saracevic 2007, Currie 2010
consultation with a librarian, subject guide, etc.); and the range of devices (e.g. mobile, iPad, etc.) that a person might be using. In our ideal, we expect that our users could navigate the full range of resources and connect to relevant articles, images, data, on any device, plus be able to find background resources (subject encyclopedia content) and personal help at any stage in the process.

Another functionality requirement comes out of our ongoing work on developing virtual libraries as physical facilities close or merge. Patrons need to be able to search subsets (generally subject centered) of our resources. For example, the Physical Sciences Online Library presents classic texts in physics, and lists of electronic journals in astronomy.

Users could easily connect to digitized full-text content (whether scanned by CUL or Google) and images (whether held in ArtStor or Luna). The capabilities of the discovery system would range from browsing to faceted searching, and the user would have a clear indication of the type of information (e.g., metadata vs. full-text) that is being presented.

One library site came closest to our ideal--North Carolina State University. The site uses Summon as a metasearch tool as well as a federated search system, Quicksearch, that cuts across the local catalog, Summon, local digitized holdings, and the library's website to present the search results from each of these sources in a coherent display. The search can be easily expanded to specialized subject databases, the Triangle Research Library Network, WorldCat, Google Scholar, and to local repositories. Features such as Autofill and "did you mean" help the user; displays present relevant information, including the location and call number in a brief record. Better integration of subject guides, reference material, and greater visibility of user help would enhance the system but overall the site was impressive.

**Gaps and Shortcomings**

We determined that the current CUL discovery system has major shortcomings that impede or complicate research. For one thing, the system lacks a single search box that would allow a user to search across all the component systems, with the results of the search being differentiated according to their source. Other problems, detailed below, include, but are not limited to, the exclusion or obscuring of many relevant available resources from search results; inefficiency in performing certain types of searches and inability to perform certain other types of searches in our current faceted catalog implementation, WorldCat Local; confusing search results due to widely varying coverage and content from providers and discrepancies among the results returned by aggregators; and search functionalities that differ across the component systems of the discovery system. These add up to a discovery system that frustrates many of our user communities, most notably, faculty (some of whom have been quite vocal about their displeasure).
Scope of discovery excludes or obscures available resources

Our current system does not allow users to discover everything to which they have access. In a major university library setting, where users have access to such a vast amount of valuable, scholarly material, whether it is freely available on the Web or accessible via the library's physical and virtual collections, this is a real problem.

As an example of scholarly content that is freely available on the Web, the International Music Score Library Project (IMSLP) is a well-loved and heavily-used digital music library, comprised of digitized editions of scores that are in the public domain, or whose creators have authorized their inclusion in IMSLP. The Cornell Music Library's Web site contains a page with links to a modest number of relevant free Web resources, including IMSLP, but since our current discovery system does not index library Web pages or other free Web sites, the user who may be unaware of IMSLP is almost never directed to this astounding resource through a search in the library gateway [the library's current discovery system]. The exceptions to this rule are cases where a catalog search yields an OCLC record that contains a note indicating that a particular edition of a score is also available online in IMSLP (see OCLC # 489215988). However, these notes, which are provided as a courtesy by some kind music cataloger at another institution, are few and far between, since it is not standard procedure in the music cataloging community to check IMSLP for availability.

Another type of freely available Web resource that is not indexed in a library gateway search is the body of Library Guides created by CUL librarians. For example, a search for "anthropology" in the "Experts" tab retrieves VIVO listings for various relevant courses, but does not retrieve the robust Library Guides that were created for those courses (http://www.library.cornell.edu/resrch/libguides/2).

Moving specifically to the WorldCat Local catalog search powered (about which numerous faculty members have complained), items that should have resulted from a given search sometimes are not included in the initial list of results, and can only be discovered via an unintuitive work-around. For example, an Advanced search for the title Carmen, limited to the musical score format and to the Cornell Library system, yields 27 results, but only the first two and one other much further down the list are relevant. Clicking on "View all editions and formats" under the first hit retrieves 57 items (many of which are not actually scores, by the way). Subsequently refining that list of 57 items using the format facet retrieves 12 scores. The ninth and tenth items in this new list of 12 are the most relevant items imaginable for a person looking for a score of the complete opera Carmen, and they are arguably more desirable than the first two results in the initial set of 27 items, since they bear much more recent publication dates. These two scores never showed up in the initial search, and there is no way users would know they existed unless it occurred to them to try expanding and then refining. However, this burden should not be placed on them when the initial search is structured in such a way that these items absolutely should be retrieved.

In addition to this problem of hidden items, WorldCat Local entails countless difficulties in searching, not the least of which are difficulty in performing a known-item search (although this may be achieved reasonably well through an Advanced search) and
inability to browse by author, subject, uniform title, or call number. Even after a well-constructed search has been executed, the user is faced with obstacles in determining the value of each result, due to various display problems, admirably documented by the Music OCLC Users Group (see http://musicoclcusers.org/WorldCatLocal20100412.pdf).

(Despite the myriad problems inherent in WorldCat Local, the ability to search the holdings of libraries worldwide is of tremendous benefit to users. If Cornell were to implement a faceted catalog other than WorldCat Local and also move to a system like that of NCSU, perhaps it would be beneficial to include WorldCat.org among the group of "more search options" outside the library.)

Confusing search results

The uneven granularity of resources returned by searches on the current CUL discovery system creates a disjointed flow from the original discovery request to the final delivery of information. Resources are often available from a multitude of suppliers with a variety of availability options.

The format and coverage of the content available for a single source can vary widely from provider to provider. An example is electronic access to the New York Times. The electronic version of the New York Times is available to CUL patrons through eight different providers, with seven different full-text date ranges. In addition to the number of options that patrons are presented for this resource, there is no guidance offered to the differences among the options (other than the sometimes redundant date ranges) such as depth of content indexed and available, for instance, or variations in the content document formatting. Plain text, HTML, or PDF documents provide different levels of access to the originally published content. The discovery layer in this example (multiple, inconsistent providers to choose among) leads to a confusing journey from discovery to access.

Another source of confusion for users is due to the large discrepancies among the results returned by the aggregators across the range of titles they include. When we provide links to aggregators (e.g. ProQuest), it's difficult to predict the level of coverage. Some aggregators provide metadata only. Others provide the article text but include only selected articles. Publishers may impose a full-text embargo for time periods which vary by provider. Some periodicals publish special editions that are not included in our electronic access. Some information available in print journals and other periodicals is excluded in the electronic iteration, e.g., charts, graphs, photos, inside and outside covers, front matter, editorial board lists, newspaper advertising, etc. In addition to the user frustration caused by these issues, these variations make it difficult to cancel print version of publications when the coverage in the electronic version is not comprehensive or current.

How search functionalities of various systems differ

Further confusing the users' experience is each vendor's interface is different. Users have to learn multiple interfaces to do their discovery and retrieval if their topic requires searching several databases. The interfaces look different, they behave differently and
they offer different sets of functions. The default search protocol can be phrase search, phrase if two words and proximity if more than two words, Boolean AND between terms, and so on. Truncation and wildcard search symbols differ. While most systems have several modes of searching (basic, advanced, command line) the default screens differ, some systems default to basic searching, some to advanced. Setting limits or filters varies. Some systems offer check boxes, some require highlighting a limit, some systems use a combination of check boxes and other interactions. The overall result of this Tower of Babel of search interfaces is patrons tend not to use any of the advanced functions of these interfaces.

- Utilize an “Everything” search box that yields results differentiated according to their respective sources/types (such as articles, books & media, etc.). The catalog interface for books & media would include facets which users can employ to refine and reorder their initial results (date of publication, format/resource types, full-text availability, language, peer reviewed, subject). (See Appendix A.3 for an example based on North Carolina State University’s system).
- Build in limiting to specific types from the default search screen: Cornell only, articles, books & media, background, etc. (Appendix A.3).
- Provide the option to search by call number.
- Provide pertinent information for each item at the results level: location, circulation status and frequency, call number, edition date, availability of full-text (including BorrowDirect and ILL), and contextual “did you mean” help.
- Incorporate visible access to personal help (via chat, texting, phone, and email) at all search and results levels.
- Expose user-selectable, subject-specific virtual library content and background resources (encyclopedia content, research guides) in context.
- Build in one-click access to the full text/full image sources across vendors and easy access to non-Cornell resources held by consortial partners and other libraries.
- Allow easy saving and exporting of results (e-mail, texting, RefWorks, etc.).
- Accommodate a wide range of user devices (such as smart phones, iPads, etc.).
- Provide secure and simple authentication and access to Cornell-only resources from on- and off-campus locations.
Appendix A.1: Specific Recommendations from Focus Groups

Following are specific recommendations based on feedback from the focus groups that were conducted.

Search

- An "Everything" search box that yields differentiated results (books, articles, applicable journals, databases, websites, etc.). See screenshot in Appendix A.3 for example (based on the North Carolina State library search screen).
- Ability to toggle between searching Cornell collections and wider world - If not from the "Everything" search box then from the catalog search. A check-box for Cornell only on the first search screen was suggested. See screenshot in Appendix A.3 for example.
- Ability to limit results to full-text availability and/or peer-reviewed from the search screen. See screenshot in Appendix A.3 for example.
- Call number searching - If not from the "Everything" box then from the catalog search
- Autofill - the search box automatically fills in text as the user types. Autofill can sometimes be overzealous so it would be important to have the capability to turn it off (PubMed is a good example).
- Book search on the first screen - to be part of "everything".

Results

- See library, status and call number on the first search results page. Users should not have to click on an individual item to find this information.
- See the date of an item is on the first results screen (the edition date, not when Cornell acquired it).
- Ability to see items that are available full text online from the initial results screen without having to click on the individual items.
- Ability to limit results to items that are available in full text.
- An improved "Did you mean..." functionality. When a user types a misspelled word in the search box, the system should provide a suggestion. The current CUL search system provides a limited version of this request - it will only provide a suggestion if there are no results, and the suggestions are sometimes wrong (i.e. type "sustaneable" in the catalog search and you will receive a "Did you mean sustainable?" suggestion).
- Ability to refine the results by adding additional keywords to the search.
- Option to see similar results.
- In catalog search results, see how many times an item has circulated.

Export

- Login to in-database accounts (to retain search results, etc.) should work with a Cornell netID and password, instead of being required to set up an account.
- Export to RefWorks directly from the results screen.
- Ability to send call number list via text message.
Remote Access

- Users have expressed frustrations that Passkey doesn't always work when they are off campus. Springer journals are problematic. Anything that can be done to improve off-campus access for users is important.
Appendix A.2: Dimensions of Discovery and Access

Type of Resource
- Indexing and Abstracting services
- Subject Specific research guides and webliographies
- Encyclopedia articles and dictionary entries
- Journal Articles
- News reporting and newspaper articles
- Articles, Essays or Chapters (edited books)
- Book-length Monographs (one subject)
- Images, including maps
- Statistics and Data
- Audio
- Video
- Other Multimedia
- Musical Scores
- Special collections materials: Manuscripts letter diaries etc
- Laws, Patents, and Government Documents
- Conference papers and Presentations
- Working Papers and Gray Literature
- People: Subject Specialists; Reference and Research Staff

Research Process
- Topic Discovery
- Gain Background Understanding
- Gain in-depth Understanding
- Information Analysis
- Citations, documentation
- Known Item

Help Availability
- Live staff, face-to-face, chat, telephone
- Email, Facebook, Twitter
- User Reviews
- Subject Guides
- FAQs

Technical
- Bandwidth
- Screen size (e.g. computer tablet, phone)
- Software (e.g. app, windows Linux)
Appendix A 3:

What this could look like

The following are screenshots of what it might look like. This is based on the design at North Carolina State University. The actual layout can vary (eg. The tabs could go along the left-hand side as the currently do on the CUL website).

Based on feedback from end-users, a “Search All” box that yields differentiated results is the highest priority.

Other high-priority items include:
The ability to limit to peer-reviewed and full-text online:
The ability to switch easily between searching Cornell only and Libraries Worldwide:

In the above example, “More Search Options” lets you choose from a drop-down menu for title, author, call number etc.

The Website tab would be the equivalent of what is currently the search box in the upper-right of the current CUL homepage.

“Background” Search results would come from online reference sources:

   
   [Full-text](#) | [3 PDF page(s)]

2. **Sustainable Agriculture**.  
   
   [Full-text](#) | [2 PDF page(s)]

   
   [Full-text with Graphics](#) | [2 PDF page(s)]

4. **Vanilla Industry**.  
   
   [Full-text](#) | [2 PDF page(s)]

The Experts search would pull in results from VIVO as it does currently.
The above example includes:

- See library, status and call number on the first search results page
- See the date of an item is on the first results screen (the edition date, not when Cornell acquired it)
- Ability to see items that are available full text online
- Refining - ability to add additional keywords to the search
- In catalog search results, see how many times an item has circulated

**Citations From Text**


Appendix B: Information Architecture (Group 3)

Explore an integrated "layer" that would bring together diverse information sources, integrate our "value added" component - tags, enrichment, etc., and support the creation of user-focused views and systems. Consider candidate technical architectures for the "layer". How would it be populated, maintained, and used? Identify sources to be integrated; identify examples of adding value; describe how this information would be accessible to the user.

This report is based on two assumptions:

1. No vendor product will provide a viable solution to connect to and index content in CUL's various local collections and value added information alongside our licensed resources. If it turns out that this assumption is wrong, and some vendor product can be used, then that may obviate the need for this layer.
2. It will not be possible, both because of permissions and the impracticality of dealing with very many sources, to build a system where we locally aggregate metadata and full-content for all licensed resources.

The combination of these two assumptions means that there will be aggregation both locally and within a vendor service. This integration layer must support effective indexing and use of the both sets of aggregated content.

Architectural options

No vendor solution will provide a viable solution to connect to and index content in CUL's various local collections alongside our licensed resources. We thus imagine that there will be a local aggregation component that aggregates data from at least our local collections and then either feeds external discovery systems and/or a local discovery system. We see the following possible configurations:

1. We aggregate content locally and then feed all information to an external system for indexing. Users see just the user interface provided by the external system and this integrates search over both local and licensed content.
2. We aggregate content locally and provide a discovery user interface over that; we rely on one or more vendors to index licensed content. Results to an initial search are presented in NCSU fashion but likely with fewer different areas.
3. We aggregate content locally and then integrate this with vendor index of licensed content using a locally implemented federated search over the vendor API. We present a single user interface to the user which integrates search over both local and licensed content. This is not a promising option because of the difficulty of integrating the rankings from independent searches.
The common thread to all of these configurations is that we require an infrastructure to aggregate local content that is to be indexed or search with externally aggregated content.

**Sources to aggregate**

**Local collections**

We should provide integrated discovery over all locally held collections, and all collections listed in the Registry of digital collections should be aggregated and indexed. The following table provides a sampling of some of the larger collections and spans the different technologies and likely aggregation methods.

General web-crawled data is not included within this effort; we are not trying to compete with Google. However, web resources identified by selectors, crawled by ARCHIVE-IT and archived under our management by the Internet Archive should be included.

Information about subject librarians will be incorporated by harvesting their VIVO profiles.

The scope will be limited to scholarly resources. Thus general web pages, general library website pages, and personal websites for librarians not in VIVO (should ensure VIVO record is good) will not be included. Luna has not been included because it is being phased out in favor of ARTstor.

**High-value local information**

In addition to local collections there are various Cornell managed resources of high-value additional information that should either be included directly (e.g. librarian profiles), used to indicate content to be included (e.g. links to resources in the (Hotel Library website ), or used as hints to ranking of otherwise included resources (e.g. a link to an arXiv article on a LibGuide, or circulation statistics for a book). Sources of such information are described here, cases where they may be used to enhance the value of the discovery and access system are discussed in the next section. The following sections list important sources of local information grouped into three classes.

**Managed systems**

There are a number of managed systems used by CUL staff and other members of the Cornell community that contain high-value local information. Access to this information will require use of APIs where available, custom interfaces or perhaps even web crawling if other facilities are not available.

- **LibGuides** - there are ~560 LibGuides maintained by CUL librarians, typically containing a few pages of content which has both static and dynamic components. Examples include:
  - American Historical Newspapers Online (LibGuide)
- **Applied and Engineering Physics** (LibGuide)
- **Course Reserves** - lists of items on reserve, along with title etc. of the course, could be extracted and this information used to influence ranking.
- **InSITE** - annotations from the Law Library. There are approximately 1600 annotations of websites which can be indexed and also used to adjust the rankings of any websites/resources included (perhaps also later used as a guide for targeted crawling/ARCHIVE-IT of the websites themselves).
- **Blackboard** - much of blackboard content is private with complex permissions. It may be possible to extract course readings or similar data.
- **Voyager** - circulation statistics, and presence of multiple copies could be used to boost the ranking of a given title.

### Library webpages

Many library websites include pages that embed resource links and so provide useful additional information. A significant number of library websites now use Drupal so it may be possible to gain API access for content, for others web crawling will be necessary. Management of which web pages should be included will require a registry of some sort and should allow individual librarians to add/change/delete entries rather relying on centralized support. There may need to be tools to help keep this information current such as alerts when pages are no longer available or period reminders to check that content is still appropriate. The following are example web pages that might be included:

- **Hotel Library** (website)
- Hotel Library's [Golf Industry Publications](#) (this has click-through tracking of usage which might make matching of resource URIs difficult)

### RSS feeds

Some lists of resources are delivered and used as RSS feeds. These may be be managed locally (e.g. CUL New Books) or in the cloud (e.g. from Delicious). In both cases the feed contain not only the list of items but also summary and tagging information. Examples include:

- **Delicious feeds**
  - [business_student's ag_econ_databases](#) (Delicious user/tag page)
  - [Agricultural Economics](#) (Delicious RSS embedded in LibGuide)
- **CUL New Books**
  - [Business @ Mann Library: New Books at Mann](#) (CUL New Books RSS and Delicious RSS embedded in Confluence page)
  - [Anthropology: Guide to Library Collections and Services](#) (CUL New Books RSS embedded in LibGuide)
Use of local data to improve discovery

Adjusting ranking

References to indexed resources from LibGuides, RSS feeds, course reserves will be used to adjust the ranking of these resources. The premise being that, all other things being equal, of two resources that match a query equally well the one that has been referred to by a local expert should be ranked more highly.

User focused views

- VIVO output (API, JSON, feeds) - value stems from the normalization and aggregation from multiple input streams
- e.g. Law Library might run some query to re-use data on their website
- might need tagging on per collection basis at least

Aggregation technologies

There are a number of types of information that a discovery aggregation layer would need to keep about the resources being described. At a minimum, these would include some form of representation for the resources themselves, metadata statements about the resources, organizational structures, annotations (both structured and unstructured), relationships among the resources, and possibly first-class representations of people who provide relationships, annotations, and other context. The goal is to capture in a useful form all that we can "know" about the resource that might aid in its discovery. There are several different technology approaches and tools that we could use to create this layer of aggregated information about resources. These break down into three broad areas, with hybrid approaches combining elements from two or more also possible. They are:

- Relational Database: In this approach, we would create appropriate table structures for all the kinds of information that we want to capture and preserve about the resources. Relational databases are a very efficient way to represent and query structured information, but there can be a high design cost for creating both the database and queries, particularly where the relationships are complex.
- XML-based representations: In this approach, the fundamental representation of the data is in XML. Fedora, and in particular the network overlay approach taken by the National Science Digital Library project, is based on XML, although it is supported by both a relational database and a semantic-web triplestore cache for fast querying and access to information. Note also that some relational databases (e.g. Oracle) now include native support for storing and querying XML.
- Semantic Web Technologies: This breaks down all the aggregated information into simple subject-predicate-object statements that are interpreted according to a defined ontology. This is the approach used by the VIVO system for creating faculty and researcher profiles, and the underlying VITRO code provides an ontology editor, querying, and display system for these triples. This is probably the most flexible approach, in that new types of statements and relationships can be added easily to extend the aggregation layer. However, efficiency can be a
significant challenge, and some kinds of structured information that is easy to represent in XML can be very complicated to represent and edit as triples.

Currently, we are strongly inclined toward using a VIVO/Semantic Web approach to creating the aggregation layer, for the following reasons:

1. We have staff with deep expertise in this approach, and we can continue to leverage the work that is going into the VIVO project.
2. By drawing on VIVO data, we already have first-class representations of librarians and other potential contributors to the aggregation layer.
3. VIVO is designed to support the incremental harvest and aggregation of information from multiple data sources.
4. Triples are a highly flexible and extensible way to make statements about resources and to assert arbitrary relationships. New information types can be easily added without affecting existing representations and queries, something that is far less true of the XML and relational approaches.
5. It is very easy to expose our own contextual information as RDF so that it can be freely reused and incorporated into other systems. Similarly, other RDF-based sources of information about our resources can be easily combined with local information, including straightforward ways to state equivalences among identifiers (e.g. author IDs).

Registry of digital collections

The current registry of digital collections was implemented to serve as a switching engine for federated/aggregated searching across local and remote databases. This goal was never realized and the registry has been used to keep track of and document local collections (by about 3 people in Metadata Services). This effort provides a good opportunity to revisit the original goal but at the same time we must revisit the technology and implementation choices for the registry.

If we are also to maintain lists of web pages and RSS feeds to extract information from then these should also be managed in a registry. This could be done either as an extension of the same registry that handles digital collections, or as a parallel service.

Current implementation and issues

The current registry uses the orphaned Ockham platform which has seen no development since 2007. The Ockham user-interface is tied to the storage schema, which is tied to a particular Java peer-to-peer technology, which is also tied to the Java OAI-PMH harvesting interfaces. Very little of this is necessary for the registry now, and simply gets in the way of customizing this for our purposes. Our current web user interface does no use this code and simply accesses the Ockham database. Any development of the registry beyond its current purpose should involve migration to a new platform.
Migration to VIVO

The preferred option for development of the registry or registries is to adopt and semantic approach using VIVO. This offers excellent integration with other library information. Some additional VIVO work would be required to add support for restricting editing privileges to certain user groups for certain subsets of the data (graph). Migration would involve conversion of the data to RDF based around and initial ontology that would be developed from the current schema.

Another options would be to migrate the back-end to a schema free database (e.g. CouchDB, MongoDB) to avoid current schema maintenance issues. This would involve writing editing facilities and implementing an OAI-PMH harvester. The front end might also be rewritten in Drupal, consistent with many other library web projects, and would require less ongoing maintenance than the current highly tailored front end.

VIVO for aggregation

Technology background

VIVO is an application built on top of a general-purpose ontology editor, content editor, and web display tool called Vitro, and also developed here at Cornell in the Library. Vitro is used on several other projects including DataStaR, by a consortium of Australian universities supported by the Australian National Data Service (ANDS metadata stores solutions), and by the Chinese Academy of Sciences for "subject knowledge environments" in biodiversity and medicine and health. There are also some early Vitro-based applications for international journal access and for keeping track of projects in CUL-IT.

So what is VIVO? VIVO extends Vitro via a much more detailed ontology to support information about researchers and their activities and research outputs at a university. VIVO also has software extensions to support a more customized interface for tasks such as adding a publication and its authors, and a navigation structure closely tied to its ontology, with major headings of People, Organizations, Research, and Events.

Any semantic discovery and access layer for general use as one part of a larger discovery system at Cornell would likely be built on top of Vitro, with an ontology developed to represent the characteristics of the resources represented in the discovery system, of the sources of that information, and of any annotations or value added information expected to be gleaned from librarians, from usage statistics, or eventually from end users directly. One principal value of the Vitro approach is that the data model for the application is not fixed – if we need to extend our background layer to support a new type of information such as relationships between information resources and common reference questions or to identify any resources that have been on reserve for a course in the past 5 years, this can be accomplished by extending the ontology rather than by having add new tables or fields to the underlying database. In practice, this is not always as simple as it sounds,
because some information should be grouped or ordered in a particular way to make sense for editing or display purposes, and some customization of the interface may be required.

One rule of thumb is that for any application that is clearly defined and will not grow significantly in scope, it will probably be a more predictable outlay of programming resources to use a traditional database approach, whether through Java servlet/JSPs, Drupal, straight PHP, or content frameworks based on Ruby. When an application is breaking ground in a new domain where the requirements are not well understood, however, Vitro offers a much-more open-ended solution that will be easier to extend based on the results from early prototypes and user feedback.

The tradeoff will often come from the need for complex editing interfaces and customized display solutions. Increasingly we are looking toward implementing these application components in Drupal, and using Vitro as a persistence layer with unique data representation and query capabilities. This has been successfully demonstrated through sites such as the CALS Impact and Experts sites (about to be merged for the CALS Communications Office) and by an earlier hybrid solution called the Graduate Education portal. Using another application to post editing changes to a Vitro application has not yet been demonstrated, but the Vitro editing layer anticipates this by accomplishing it's final changes to the database via sets of additions and retractions, which could just as well be done from an external source providing that appropriate authentication could be put in place.

Under the covers, Vitro stores its data through a semantic application framework called Jena, and Jena manages persistence of Vitro data in MySQL or other standard relational databases such as Oracle or PostgreSQL. Vitro sees the data and works with it only in the form of triples, or three-part statements linking a subject, a predicate relationship called a property, and an object. The object of the statement is sometimes a simple text, number, or date value and other times another "individual" – another entity that can itself be the subject of other statements. These statements are governed by an ontology, a structure of types ("classes") and relationships ("properties") expressed in the Web Ontology Language referred to as OWL. Much more information about OWL and about the Resource Description Framework or RDF data model used to represent these triples can be found in Wikipedia or on the World Wide Web Consortium or W3C site.

**Data vs. Metadata**

It's sometimes very confusing to distinguish what we mean by metadata from what we refer to as data, and this distinction is made no clearer in the semantic world, where the same technologies are in wide use for storing the individual record-level facts of data and for managing metadata about information resources stored outside the semantic environment in a more traditional repository. The DataStaR (short for data staging repository) project has been created to provide a venue for researchers to upload datasets, document them, share them with colleagues, and then (optionally) submit them to a repository intended to hold them over the long term. In this case, Vitro serves only as a
metadata store that references the data files uploaded to a Fedora repository; when the dataset is submitted for publication, it would be zipped into a submission package or SIP appropriate to its destination repository along with the metadata from DataStaR.

This pattern is likely to follow with any semantic discovery layer – it is likely to consist of metadata about a diverse collection of information resources along with references to the full text of the resources, to images, databases, video clips, and even other websites stored elsewhere, be it the Internet Archive, the Hathi Trust, or a local collection.

Provenance

One likely requirement for an integrated discovery support layer will be to keep track of where any of the information stored has come from. In an information architecture based on simple triple statements, it's not possible to add another "field" to the basic statement without altering the fundamental paradigm, so a different approach has to be devised. Instead of adding another column or field to a row of data, where one row represents all the information about any individual entity or object being stored, RDF adds additional statements about the same individual, and the software pulls in all the statements with that individual as a subject. This works well for managing lots of different statements about an individual, but still doesn't solve the problem of knowing where each statement came from. This is handled by storing each statement as a member of a collection of statements known as a graph – "an abstract data type representing relationships or connections" (graph data structure). The graph itself can be described by statements identifying its source, date and time of last update, and other information relevant to establishing the provenance of any statement.

Vitro has recently (with VIVO 1.2) been modified to use the Jena graph structure as its preferred storage library, and further development is slated for this functionality to support selective operations on data based on its source as well as identification of the source itself.

Ontologies

One significant task for developing a semantic layer will be identifying candidate ontologies for bibliographic data. VIVO has adopted most of the Bibliographic Ontology to represent publications themselves, but this will need to be supplemented by additional ontology elements (classes and properties) representing:

- citation (e.g., CITO)
- provenance, authoring and versioning (SWAN PAV)
- annotation (e.g., AO)
- scientific discourse (SWAN discourse relationships)
- online communities (SIOC)

Another ontology activity will involve mapping from the composite internal semantic representation to Dublin Core and other library standards.
- Expressing Dublin Core metadata in RDF (DC-RDF)
- functional requirements for bibliographic records (FRBR) (expression of core FRBR concepts in RDF)

**Consistent identification of same objects**

*(so join works in combined graph)*

One of the significant challenges for any background integration layer, no matter the technology, is that library resources are available from many sources – we may have multiple ways of finding out about the same book, journal, article, or derivative work. We will have to develop

**Incremental update for input collections**

The important thing here is that to do incremental updates without string matching, we either need to have an external identifier in the source data or be able to add the VIVO URI created for that resource to the source data.

We have spent considerable time and effort devising incremental update for VIVO feeds here at Cornell, so I think we understand at least many of the issues.

**How to support incremental update of output feed**

(back of envelope in meeting says that at 0.1s/page output (say XML for one item) that means a dedicated machine could do 6million pages/week so for the sort of collection size we imagine (~5million?) we could refresh a cache of all XML output on a weekly basis and do incremental feeds to a vendor system from that)

**How to feed data out in correct format**

Data for an item can be assembled, serialized and styled in XML or JSON in a fashion mirroring the creation of HTML display pages. This could be done in any format necessary for the indexing system. All data from the aggregation layer should be accessible via a public API, so that virtual libraries, websites, and other tools can use this data in arbitrary and novel ways.

**Population and maintenance**

The key steps in populating this layer will be:

1. Identify all sources to include, add information to the registry, which will be used to coordinate the aggregation of these sources.
2. As necessary, develop custom tools to harvest from collections that do not support standard APIs.
3. Harvest from all sources, understand update schedules, monitor.
4. Develop transforms to convert and normalize data from different sources.

We should not underestimate the effort required to harvest and normalize data from multiple sources (cf. NSDL experience http://arxiv.org/abs/cs/0601125). Ongoing effort will be required to watch incremental processes, tailor to updates, and to add support for new collections as they are developed.

Some of the local collections to be aggregated will be ingested into the CUL Archival Repository (CULAR). Aggregation via CULAR offers a uniform interface which may avoid building custom harvesting solutions for different platforms. It will not avoid metadata normalization issues however, because CULAR will simply ingest the native metadata.

**Exposing data in the aggregation layer**

The primary way in which information collected in the aggregation layer will be exposed to users is via the search service supporting search over both this and remotely indexed content. This would be via local search or vendor search depending on the architecture adopted. The layer will also facilitate user access in several other ways:

- Portions of collections can be exposed to web crawlers and thus indexed in general web search engines. This is useful where the source collections are not already indexed.
- Portions of collections can be exposed for direct user access via VIVO's presentation and browse facilities.
- Other tools, general or specialized, might be created based on the normalized collection.

**Recommendations**

- We should create an integration layer based on RDF data using the VIVO platform.
- Work to create an integration layer will require a registry of collections to aggregate. We again recommend the VIVO platform. This should replace the current registry of digital collections which runs on an obsolete and unsupported platform.
- We should, where possible, reduce the number of different silos and separate collections we maintain.
- When considering new technologies for collections we should ensure that they provide APIs for metadata and full-content harvesting on an incremental basis to support aggregation and indexing
Appendix C: Examining Internal Processes at CUL (Group 4)

**CHARGE:**

New types of resources, specifically those that are characterized by specialized interfaces, are constantly becoming available. What is the internal process by which these are integrated into the CUL discovery and access landscape? Describe relationships between existing CUL groups and models, including the current decision process for funding and supporting new discovery and access related services. Make recommendations for a more responsive approach to supporting and upgrading our discovery and access systems, including a more effective solution for managing trouble tickets and enhancement requests.

Our team gathered three sets of data for the analysis. The first is the charges from some 12 CUL departments, committees or groups responsible in one way or another for our discovery and access systems; we do not know if this list is comprehensive. The second data set is a collection of different case studies that illustrate various known problems created as a result of the way discovery and access is currently implemented across CUL. The third data set is a review of a sample of high level internal project documents written over the past 10 years.

From this data, several themes emerge:
- There is a need for clearer roles and responsibilities for staff working on projects,
- maintenance for ongoing services and integration with CUL's existing infrastructure are not properly resourced, and
- CUL's organizational culture minimizes the commitments and costs of ongoing enhancement and maintenance and for the most part provisions only for development and other up-front project costs.

**Introduction**

The discovery and access landscape is evolving at a rapid pace, and the 2010-2011 Discovery and Access Committee has been tasked with "develop[ing] recommendations to LibExec for processes and systems to dramatically improve discovery of and access to digital and physical information resources relevant to the scholarly and creative needs of Cornell faculty, staff, and students. They will seek to leverage recent developments in discovery systems and to address a variety of issues with existing approaches" (Cornell University Library, 2010a). Much attention has been given to the complexity of emerging discovery and access systems; however, institutionally, CUL has been navigating how best to provide comprehensive and seamless access to heterogeneous digital content for quite some time.

The focus of this subgroup is to consider the "relationships between existing CUL groups and models, including the current processes for supporting new discovery and access
related services" (Cornell University Library, 2010b). To explore these relationships, this report will first provide an overview of the landscape of departments, committees, and groups responsible for access and discovery systems. It will describe the consequences of an organization that is constantly undergoing structural change. The next section of the report considers several examples of complex systems and services, describing the problems caused by the way discovery and access is currently implemented across CUL. This is followed by a close examination of the nature of project-based work at CUL. In these subsections, several key themes emerge. There is a need for clearer roles and responsibilities for staff working on projects, maintenance for ongoing services and integration with CUL's existing infrastructure are not properly resourced, and CUL's organizational culture minimizes the commitments and costs of ongoing enhancement and maintenance and for the most part provisions only for development and other up-front project costs. This creates a gap between these explicit public goals and the processes in place in the organization to reach these goals.

The report presents several recommendations to improve the effectiveness of discovery and access endeavors in CUL. First, the library ought to support flexibility to support and develop, when necessary, grass-roots, light-weight solutions for systems and services that do not fit any existing organizational model. Next, it will recommend that CUL develop a culture of project management, and recommend various pathways to achieving that goal. Finally, it will make recommendations for a more responsive approach to supporting and upgrading our discovery and access systems, including a more effective solution for managing trouble tickets and enhancement requests.

**Landscape of groups**

CUL has a large number of groups with distinct responsibilities relating to various aspects of discovery and access. (See Appendix C.1 for a list of relevant groups and their respective charges.) These can range from formal departments, long-standing committees, temporary task forces, and others. Taken collectively, the responsibilities of these groups often overlap significantly. This has implications for effective communication, as it can be difficult to know when different groups are tackling similar issues. There is no one place to monitor all relevant ongoing work. For all of the overlap that does exist between the groups, there are still sizable gaps present. There are several resources and services that do not fit neatly into a specific category and so cannot be logically assigned to any one group. One example of this is the bX Recommender, a resource that has currently not been purchased by CUL because it fits into the purview of no existing department or group. Another example is testing mobile applications that deliver library content: multiple groups are clearly involved with mobile applications, yet none have the resources to sufficiently handle testing new applications. Both of these examples will be explained in more detail.

In addition, CUL departments, committees, and groups are in constant flux due to staff turnover and repeated organizational reorganization. Change can keep the organization responsive, but it does come at a high cost: this “churn” can have a deleterious effect on the ability to sustain efforts over time, since there is a lack of consistency from year to year. Constant change creates uncertainty about individuals’ roles and responsibilities.
Consider the Digital Library and Information Technology unit (DLIT), which has had no less than five structural reorganizations in recent years. It can be argued that this state of constant flux affects the organization’s ability to provide resources for ongoing services and administer large, longer term projects. This can be observed in several of the projects described in subsequent subsections: constant organizational change often left critical support functions unassigned.

Case Studies
Overview of issues
Enhancing the discovery and access of library materials, with specific attention to the infrastructure that is needed to support digital resources has been a priority for Cornell University Library for at least nine years. The 2002-2007 CUL Priority Objectives document states that CUL is to "implement an integrated and methodological framework Library-wide, providing users with an integrated approach to discovery and use of Library resources" and to "pursue seamless linking between e-resources" (Cornell University Library, 2002). The 2007-2010 Strategic Plan notes that CUL should position itself to "unify access to library content across all Cornell campuses" and "simplify resource discovery and access to materials, regardless of location or ownership" (Cornell University Library, 2007). The most current Strategic Plan asserts that CUL should "dramatically improve systems for finding and using scholarly resources" (Cornell University Library, 2011). The notion that it is not trivial to link together the varied digital resources needed to support teaching, learning, and scholarly activity has been acknowledged.

Over the course of the past decade, there have been resources and services that, for one reason or another, have been difficult to fit into the library's existing infrastructure. In examining these cases, some common themes emerge: there is uncertainty about roles and responsibilities of individuals and groups, individuals and groups often do not have the appropriate resources or institutional support to meet their goals, and specifically integration with CUL’s existing infrastructure is not resourced at the level it needs to fully succeed. In short, much of the (often project-focused) work lacks the conditions critical for its success. In considering the previous findings from the groups in the 2010-2011 Discovery and Access Committee, it is important to be cognizant of the challenges that the library has faced when planning for and implementing discovery and access systems. This section will illustrate some of the issues facing several systems and services – the bX Recommender, SciFinder, Ares, CUL’s “new website request group”, and CUL's high-risk digital collections.

bX Recommender
The bX Recommender, developed by Ex Libris and researchers at Los Alamos National Laboratory, is a service that applies collaborative filtering to scholarly articles. The service would be plugged into Cornell's branded link resolver service, Get It!, to display relevant articles alongside the one that is sent by the user to the link resolver. Originally, the service was only available for institutions using SFX, but in the fall of 2010, it became available to non-SFX customers. Technically, the service is simple to implement, requiring only a line of JavaScript inserted in CUL's existing link resolver. The price for
this service, $3000 annually, is relatively inexpensive; however, it has not been purchased.

Essentially, the problem boils down to the fact that the library has established budgets and mature workflows for purchasing content, but not for purchasing services (it should be noted that this problem was identified in the 2007 Web Vision report). The bX Recommender service has high potential value relative to its cost, yet there is no group currently in CUL that has clear authority over this purchase. It may be the case that no one particular group has clear jurisdiction here because the bX Recommender is qualitatively different from the resources the library has historically purchased. CUL needs a process for handling these types of requests. Additionally, relevant groups and individuals ought to have the support to make decisions and supply resources for new systems and services that defy classification into CUL's existing categories. The organization needs to balance the need to delineate clear roles and responsibilities for systems and services while also maintaining enough flexibility in order to respond to ambiguity. The bX Recommender case study provides a compelling argument for this balance.

SciFinder
SciFinder is one of the key resources for the Chemistry communities at Cornell. It indexes over ten thousand journals, patents, eBooks, and other data sources, and has millions of bibliographic records. All of the information within it has been curated by PhD chemists, and so it is highly valued for its accuracy and trustworthiness. Currently, its funding is managed by the DRC, and it was one of the first library resources to have multiple funding sources from several other science libraries. Today SciFinder is a web resource, but historically it was software (specifically, two installers) that users installed on their local machines. The software provided the authentication to SciFinder's servers (to verify they originated from an institution subscribing to the resource) and handled the complex searching interface (which includes a graphical substructure search). It is, however, a complex resource for both its content and features and this complexity presented a number of challenges to providing access to it (L. Solla, personal communication, January 2011).

Providing robust access to the installation files for the client version of SciFinder was complicated by both the lack of necessary resources to directly support it and uncertainty over responsibilities for all involved in managing it. Integrating the resource into CUL’s existing infrastructure was also a challenge: since the catalog could not accommodate serving up software, the Voyager record for SciFinder pointed out to an external page that included links to the installers and related documentation. Each year, CAS (the company that puts out SciFinder) would update the installer, and the external page subsequently needed to be updated to include those files. Initially, the Chemistry Librarian did not have direct access to the server that hosted the files. The contact person responsible for the server was not consistent from year to year, causing CUL's SciFinder page to be out of date while the situation was straightened out. After several years, the Physical Sciences Library was given its own resources – access to space on the server that hosted its unit website – to manage providing the installation files and documentation.
Extensive technical support for library resources is often not needed: the majority of the networked electronic resources that CUL provides can be accessed with simply a web browser. Subsequently, there was initially minimal organizational structure to support SciFinder due to its complexity. Since scant documentation from CAS was available at the time, it was necessary to run through the SciFinder installation process in order to create complete support documentation. Relatively few library staff had administrator privileges on their workstations, making this extremely difficult to do. Until 2004, there were no library-managed machines available to test out the Mac installation. The support network for SciFinder grew very slowly over the years. The Chemistry Librarian was the main point person for installation help and worked with several technical support teams over time to grow the support for its users, including Desktop Services and Electronic Resource & Serials Management (ERSM) in the Library, and technical support in the various departments. Essentially, it is crucial for the Library to provide service providers the conditions critical for their success; in the case of SciFinder (as well as the other discovery and access tools and services outlined in this section), this includes the organizational flexibility to quickly and responsively support complex discovery and access systems.

*Ares*

The Ares software is a system for managing course reserves. It is “designed to consolidate and streamline library reserve processing across libraries and provide the greatest ease of use and functionality for faculty, students, and reserve staff” (Cornell University Library, 2009). While Ares has indeed done much of what it promised and which the group chartered to examine its merits hoped to do, it nevertheless has proven susceptible to several problems due to lack of integration with other CUL systems (T. Trutt, personal communication, January 2011). Almost every implementation issue with Ares can be traced back to these issues: importing data from an existing CUL system or automating another CUL system. Without the necessary desktop tools, reserves processing time is actually increased. It has been possible to address some of the issues through the desktop tools, as well as the rewrite of the Ares web interfaces, but there are still considerable “holes” in the system needing to be repaired.

Some other issues with Ares stem from CUL’s specific implementation of Ares. To give an example, Syracuse University uses Ares and has few problems with it interfacing with Blackboard. This is because they use Blackboard exclusively: all courses are automatically created as a Blackboard course and, simultaneously, an Ares course via a register office feed. Since Cornell does not follow this pattern we still have connection issues with Blackboard. Ares was meant to be used either has a plugin to course management system or as a standalone system. CUL is using a combination of both, resulting in the difficulties with Blackboard integration.

The Ares implementation group hopes to rectify some of these problems within the near future. The central fact of Ares' lack of integration with other systems remains, particularly since some of the tasks it might normally be ideal for are already being done...
Appendix C

by Blackboard, a CIT service which CUL has no control over. Ares can be taken to represent one of the essential tradeoffs of existing technology: By choosing off-the-shelf software and adapting it to our existing environment, we save enormous time and money, but then have to be willing to accept the imperfect fit of such solutions. This is likely not a temporary problem but an axiom of technology that CUL should acknowledge.

CUL’s “New website request group”

CUL's "new website request group" grew out of an emerging need to develop "a more effective collaboration and decision-making process for Web based projects" between Digital Library Information Technologies (DLIT) and Library Communications (LC), who both provided Web development services. The service was originally proposed as a collaboration between the two departments. Its primary goals included having a collective decision making process to ensure that projects were appropriately prioritized, staffed, and guided through all stages of the Web site life cycle. It also strove to ensure that staff members were flexibly assigned to projects to achieve optimal workloads and effective use of skills. Emphasis would be placed on the initial phases of projects, making general recommendations for project implementation and longer term responsibilities. The group originally included four representatives, two each from DLIT and LC. This group managed a web form to track project requests, made the decision to undertake projects (after consultation with requestors), help launch projects, and pass them off to their eventual project teams.

The group sought to create a general structure for handling new project requests that was flexible enough to adapt to individual project needs, yet able to guide how decisions were made by providing consistent expectations for all involved. The group selected two concrete tools in order to meet this goal: these tools were Oya Rieger's "Life Cycle Management of Web Sites" and the "One Page Project Manager" template for planning a project and reporting its progress. The group also developed "indicators" to help guide communication with clients about the technical platforms available at the time were most appropriate in different situations. This was especially helpful to educate clients about the organizational impact of these decisions, particularly when there was no cost recovery involved. The group also developed guidelines to determine if a project be done on a cost recovery basis.

In evaluating the new service six months out, it emerged that the group's model worked very well at the outset of a project; however, as projects progressed, timelines and deliverables were likely to be compromised to account for newer, competing organizational priorities. This resulted in tension with the client. The initial charge and the resulting process recommendations of the group reflect this bias. This issue is not confined to the new website group, and can also be seen with the management of CUL's digital collections (addressed later in this section).

The service underscored the need for "more explicit project management" – as roles, responsibilities, and time commitments for those involved in the projects are uncertain when project management is not made a top priority. This lack of explicit project management created a situation where individuals were not given enough support to
make the decisions that directly affected the success of the project. The group was not sufficiently able to negotiate time commitments with clients or to "say no" to a project, with a lack of enforcement to the "appeals" process. In other cases, the project plan itself did not seem adequate to manage expectations, lacking an explicit communication plan with the client. The group was, at times, pressured to accept demands without the adequate resources, and then the project groups doing the implementation were ultimately left to try to sort out incompatible priorities and requirements for themselves. This created an environment in which the client could never be satisfied with the work of the project team and it was, at times, nearly impossible to provide a robust service.

CUL's Digital Collections

CUL has been digitizing content for over fifteen years. Many of the issues that have affected the other discovery and access resources outlined in this section also apply to CUL's homegrown digital collections. A significant number of digital collections fall into a "high-risk" category, as there is no strategy or funding in place to migrate and preserve the content once the delivery platform is no longer viable. It is often unclear whose responsibility it is to monitor the content to ensure ongoing access and functionality. Matters are further complicated by the historical lack of a preservation system at Cornell, making it extremely difficult to recreate content in new platforms. Many of the collections are of high value to the library and its users and generate significant traffic and interest globally, so it is crucial to ensure they persist. "High-risk" collections can fall into a number of categories. These categories include collections that are:

- delivered and/or stored in homegrown or non-standardized mechanisms that are hard to support long-term (KMODDL, Gloopad);
- collections where the original images may no longer be available (Fuertes, Lafayette);
- collections with data in non-standardized formats (RMC exhibitions);
- and collections with no further funding.

These "high-risk" collections require considerable resources to maintain over time. For example, in 2006, CUL decided to put the Hill Ornithology Beautiful Birds Collection, a database of images created in the late 1990's, into ARTstor. To do this, staff had to spend a significant amount of time tracking down original images (and not all were found), migrating the images from a Kodak proprietary format to TIFF, exporting the database from the website, and performing further normalizing tasks to fit the adhere to ARTstor standards. While this might have been possible for this collection, it is not scalable across all of the "high-risk" content without allocating much more staffing and resources. Without a mandate from the Library to migrate and update these collections, there is uncertainty about who in the Library ought to address this issue and what their responsibilities would be. Subsequently, these collections are susceptible to gradual demise and high error rates, resulting in the potential loss of library assets and unhappy patrons. It is crucial to develop an organizational strategy for curating (migrating or de-accessioning) the content and provide the necessary resources to achieve this goal. Appendix C.2 identifies the collections that currently fall into this category, accompanied by usage statistics.
**Project-based work**
Projects are, by their nature, designed to create change in organizations: if successful, projects are incorporated into an organization's service portfolio. Much of what CUL does to support discovery and access of resources for its patrons is rooted in discrete, project-based work. As part of the 2010-2011 Discovery and Access Committee, a workgroup led by Adam Chandler and Adam Smith surveyed past CUL reports related to discovery and access to identify conditions that correlate with successful projects. The reports consulted were:

- CUL Priority Objectives (2002-2007)
- Usability Studies at Cornell University Library: New Model Recommendation (2008)
- Issues in Discovery and Access (2009)
- CUL IT Strategies and Priorities (2009)
- Ongoing Development of CUL Web site and Web services: Proposed Solutions (2009)
- Recommendations of the WebFeat Transition Team to its four sponsoring AULs (2010)

The following general conditions were found to correlate with increased likelihood of project success:

- the project is initiated by an AUL,
- the project is small in scope,
- the project has a specific plan,
- the project has concrete goals with identified criteria for success,
- the project has external deadlines imposed upon it, and
- the transition from project to a supported service is articulated up front.

Initiatives in CUL often have no identified sponsor - someone officially committed to the project that has the authority to make decisions on its behalf - and so, the implementation team is left to figure out its organizational priority, budget and staffing constraints, and goals. By definition, this is not a CUL project, and the effort will likely fail. As evidenced in the case studies, uncertainty about roles and responsibilities negatively affects the quality of service and support.

Even on CUL projects initiated by an AUL, that AUL may not fulfill her responsibilities as project sponsor. For example, when a new, competing project emerges, priorities may not be revisited. When an AUL does not acknowledge resources allocated to one project limit the resources available to all other projects, separate implementation teams must sort this out amongst themselves. For example, work on the Physical Sciences Library site went over the budgeted time due to an increase in scope. This pushed back the start of the Management Library site, which had its own increase in scope, which then delayed the start of the Olin/Uris Library site.
In the survey of CUL reports conducted by the working group, it emerged that the WebFeat Transition Team was quite successful. The project had very concrete, time bound goals, the implementation team's recommendations were actionable and measurable, and four AULs sponsored the project.

**Recommendations**

- Provide for and resource grass-roots solutions
- Build a culture of project management in the Library
- Project-based work needs AUL sponsorship,
- Every project should create results-oriented SMART goals
- Integrate an enterprise wide ticket tracking system into discovery and access operations
- Final project-based recommendations

Based on the evidence presented in the case studies and examining the nature of project-based work, this subgroup makes the following recommendations. The first recommendation is intended for work outside of formal projects, while the following recommendations address project-based work in CUL.

**Detailed descriptions for each recommendation**

**Provide for and resource grass-roots solutions**

This section presents the argument that staff members need opportunities to implement lightweight, creative solutions to problems outside of project-based work, especially when addressing issues for resources that do not have a clear niche within the library system. It will outline two examples of successful, flexible responses to emerging needs: assessing new mobile resources and organizing reference services for visual resources. The "mobile testers group" started as a series of conversations between the E-Resources team and Library Outside the Library. The E-Resources team had started to receive notifications of newly available mobile platforms from vendors. Initially, the E-Resources team kept a local list of these platforms, but was not advertising them to library patrons. The Library Outside the Library team had just released its CULite suite, including a mobile version of the main CUL webpage with a curated list of library-related mobile resources. There was clearly overlap between the two groups, though several key issues quickly emerged. How could the library test and support multiple mobile applications developed for a range of mobile platforms? Who would be responsible for vetting newly available mobile applications and mobile-friendly websites? Neither group was equipped to handle these tasks alone. As of January 2011, no CUL group formally included responsibilities for library-related mobile resources. Communication between groups was therefore essential here: Adam Chandler (acting as representative for E-Resources) contacted the current chair of Library Outside the Library, Dianne Dietrich, to tackle the issue.
Adam Chandler and Dianne Dietrich put out a call to CU-Lib asking staff in the library to identify themselves if they had mobile devices and were willing to test out new mobile applications. The respondents were added to a listserv called cul-mobile-testers. When a new mobile resource is made available, a message is sent out to cul-mobile-testers. Volunteers from that group test out the application, and provide feedback to the group. If the application appears to function well, the Library Outside the Library will include a link to the application on its list of mobile resources. This grass-roots level collaboration responded to a need to address these issues without a formal framework (within existing committees and CUL groups) to do so.

Another example of library employees coming together to fill service gaps occurred when the Knight Visual Resource Center closed abruptly in 2009. At the time, KVRC was providing support to multiple faculty members and students in the area of digitization, visual resources, and equipment lending. A team was put in place to mitigate the fallout of the closure, consisting of representatives from Public Services, Fine Arts Library, and DCAPS. Through analysis of the existing infrastructure, it was determined that the library needed to do the following in order to successfully transition services away from KVRC:

- revamp library website so as to provide more clear guidance to patrons seeking support in visual resources (with a paper brochure corollary),
- establish a listserv that would connect users to librarians capable of responding to patron requests within 24 hours, either online or via a one on one meeting to review user needs, and
- provide digitization services to faculty at little or no cost (or with funding from the Colleges).

With the support of AUL’s Rieger and McCue, all of these tasks were accomplished in a relatively short time-frame – within three months. As a result, CUL now has streamlined services, a more cost-effective workflow, and an expanded Visual Resource Working Group consisting of members from across campus.

**CUL should build a culture of project management**

To increase the likelihood of success across the life cycle of future projects, more should be done to develop a culture of project management in CUL. For a project-heavy organization like the library, establishing a project management office should be considered to develop project management expertise in the organization and recommend planning and reporting standards, among others. Project managers should be properly supported to fulfill their obligations, which might include PMP certification, Scrum Master training and mentor-ship within the organization. Specific tools and artifacts that a project manager should be responsible for might include:

- An overall description of the project, such as a Statement of Work (SOW) or A3 report, that would include the project's SMART goals, foreseeable risks (and possible responses), and a communication plan that specifies what information each stakeholder needs and how that information will be delivered to them.
• A Work Breakdown Structure (WBS), created with the help of the implementation team, to identify and estimate activities, and anticipate risks associated with each.
• A basic project plan that identifies activities/tasks, milestones and a timeline, such as a Gantt chart or One Page Project Manager (OPPM). The OPPM has the advantage of aligning activities with overall goals as well as the people responsible for each activity.
• As issues arise, a report such as the A3 that quickly communicates each issue, with explicit options and trade-offs necessary for the sponsor to make an informed decision.
• Use of a ticket system

These tools and artifacts are meant to foster communication and increase accountability within the organization. The quality of this communication should be the goal, not the specific form of that communication, although standardizing on an approach will help develop expertise in the organization across projects. In particular, the lack of periodic follow-up reports was glaring in the survey done for this report. On each project, the project manager should report at least one more time to the Library Executive Group (LEG) on ongoing progress against measurable goals, to ensure accountability in the organization. Only the minimum amount of planning, documenting and reporting necessary to optimize communication should be done, and teams should resist the tendency within CUL to produce reports for their own sake.

**Project-based work needs AUL sponsorship**

Every project must have an AUL as a sponsor before it is considered an official project and staff time is spent on it. The project sponsor has the following responsibilities:

• negotiate among the AULs for the priority of the project relative to other projects,
• allocate time, staff and budget, and
• establish specific, measurable goals for the project that can be used to judge when the project is complete, and whether the project is a success.

A proposed initiative should not be considered an official project until the project sponsor has been identified and her responsibilities completed. Obviously, these responsibilities must be carried out with the help of the stakeholders, the project manager, and the project implementation team, and has implications throughout the project that will be explored in the remainder of this report. The project sponsor also has ongoing responsibilities throughout the project, to re-negotiate all of the above in response to newly emerging projects, and more generally, to make decisions and resolve issues that arise. This is an important strategy for reducing uncertainty, ensuring that staff involved has clear responsibilities, even as they might shift over time.

**Every project should create results-oriented SMART goals.**

Many CUL projects have vague and poorly understood goals, or no recognizable goals. In these cases, the proposed goals are actually statements of CUL’s values, or more of an
elaborate mission statement, rather than something that can be measured and evaluated. Some implications of lack of goals include:

- a lack of consensus about what the project is trying to accomplish,
- mis-communication between the implementation team, sponsors and stakeholders, and
- disagreement about when the project is complete, and whether or not it was successful.

As a result, CUL projects often focus on carrying out activities rather than delivering results. By contrast, the CUL Priority Objectives report lists goals that are not specific or actionable. These "goals" are actually descriptions of activities centered on important values within the library. The scope of these activities is indeterminate and there is no proposed plan for how to implement them. To become more results-oriented, projects should define SMART goals, or goals that are:

- specific,
- measurable,
- actionable,
- realistic (but challenging), and
- time bound.

Such goals should be clear to everyone involved, and the importance of the goal for the project should be understood by everyone. Ideally, goals should also be aligned with each implementation team member's professional goals and responsibilities and become part of our review and reward infrastructure. Since the project sponsor will be negotiating with other project sponsors on behalf of their projects, they should be prepared to explain how the goals of the project align with the overall strategy of the library. Goals should also have measurable value to our patrons. SMART goals define the criteria for success for a project, and as a result, the implementation team and the sponsor(s) must not only decide how to satisfy these goals, but also how to measure progress made against them, and how that progress will be reported.

**CUL should integrate an enterprise wide ticket tracking system into discovery and access operations**

A brief survey of the organization's operations reveals an array of different methods used across CUL for ingesting and tracking support for discovery and access services. In our review of reports there is no specific discussion of how ongoing operation support and enhancements will be managed and tracked. Consequently, each project implementation group selects its own preferred tool for ongoing support, which means we have an array of different tools in production across CUL today, presented in the following table.
Email lists | Project Management / Ticket Systems | Other software
---|---|---
Libgateway-1 | Jira | Ockham (CUL Digital Registry)
Libit-1 | Sourceforce | DabbleDB
Erlm-1 | Mantis (for energ tracking) | Zoho
Form on website for new website requests | Basecamp | Excel
Confluence | Wordpress

This is obviously an inefficient way of doing business. One strategy to improve the situation would be to mandate use of one preferred system to track operation support and enhancement tickets for all new projects. Once this base system is established, new project teams wouldn’t be burdened with figuring out which system to use. Ideally, the same ticket system would be used in the implementation phase of the service life cycle, to enable more seamless transition from project to service. An enterprise wide ticket system that transcends all of our operations would significantly increase accountability.

**Final project-based recommendations**

There are other cultural shifts within CUL that should occur for the sake of implementing better projects and services. When new initiatives are considered, existing gaps in the ability of the organization to successfully carry out that initiative need to be identified and addressed. Also, management should recognize and plan for the effects that day-to-day operational demands necessary to support CUL’s service infrastructure have on resources intended for new projects and services. In general, CUL is reactive to changes in its environment, and what begins as a high priority initiative can quickly become reprioritized over time as CUL reacts to new pressures in the environment. It is much less common for a project to be started as the result of a pro-active strategy to improve existing services.

As suggested elsewhere in this report, just as there is a project life cycle, once the project is completed successfully, it may become an ongoing service, and services have their own distinct life cycle demands and best practices. And while CUL lacks many formal project management skills, knowledge of optimal service life cycle management is an even more glaring shortcoming. A possible solution to fill this gap may be found in ITIL (Information Technology Infrastructure Library) and practical applications of ITIL, such as Visible Ops. ITIL is a set of customizable, process-model based best practices for controlling and managing operations and services. The goal of ITIL and Visible Ops is to create an environment with consistently high quality services that are measurable, auditable, and constantly improved upon. Key members of CUL-IT should be trained in these practices and adapt them to the CUL environment for improved service support.

**References**
Appendix C

Appendix C.1

**CUL Committees related to discovery and access systems**

*CUL Current Awareness team:* The Current Awareness Team is responsible for identifying and promoting current awareness strategies and tools that can help Cornell community members stay up-to-date in their research fields of interest. The Team will maintain and develop the Current Awareness section of the CUL Web Site. The Team will evaluate current awareness tools, including those that are locally developed (e.g. MyContents; New Books) and external (e.g. TicTocs; Connotea) to CUL. The Current Awareness Team will collaborate with members of other related initiatives as needed (e.g. Get It! ; Library Outside the Library; Citation Management).

*CUL Information Technology and CUL Information Technology Executive committees*

- Mann Library Information Technology Services
- Library Systems and Discovery Services
- Web programming
- Desktop Services
- Mann Library Information Technology Services Operations
- Software Development and Integration
- Repository Architecture/ Services
- Contracted Services Management

*CUL Public Computing Advisory Committee*

Provide direction regarding the design and implementation of public computing throughout CUL. Advise on services being supported for users and the consistent presentation of those services library-wide. Monitor security measures to be employed on public workstations, balancing them against privacy and access considerations. Serve as a communications link for feedback and announcements between technical staff supporting public computing and library public services staff. Contribute to overall public services priorities through specific tasks as assigned by PSEC.

*CUL Reference & Outreach Forum*  
([https://confluence.cornell.edu/display/refoutreach](https://confluence.cornell.edu/display/refoutreach))

Provide leadership in maintaining a library-wide commitment to effective reference and outreach functions. Working closely with other Public Services Executive Committee (PSEC) committees, serve as an information-sharing venue for reference and outreach best practices and activities. Support the work of and provide input to sub-committees, task forces, and working groups with significant commitments to reference and outreach functions. Establish new sub-committees, task forces and working groups as needed to work on special initiatives. Maintain ongoing communication with Cornell entities that impact reference and outreach functions. Support and expand outreach functions to include areas identified as CUL priorities. Monitor and promote professional developments in reference and outreach; advise PSEC on appropriate reference and outreach strategies and methods. Contribute to overall public services priorities through specific tasks as assigned by PSEC.
Library-Outside-the-Library (LOL)
(https://labs.library.cornell.edu/)
LOL finds, configures, distributes and assesses electronic tools and services that take the library to its users, wherever they may be, with an emphasis on non-library spaces. We are not afraid of embracing trends - after all, this is where many of our patrons are at the moment. Includes: -CUL Labs CUL Labs is less of a group than a repository of projects, especially those developed by the Library-Outside-the-Library group. Aimed at the public, including users outside Cornell, it promotes projects that more widely publicize and make available the library's resources, especially resources specific to Cornell.

Mobile testers group
The Mobile Testers Group is a loosely organized effort mostly consisting of a listserve of individuals known to own and use mobile devices (esp. smartphones, tablets, and e-book readers). Individuals may submit newly announced resources (apps, mobile sites) and ask individuals in the group to evaluate the performance and appearance of the resources.

Mann Virtual Library Task Force
The Task Force will develop a plan for providing enhanced access to electronic resources and services for specialized communities of users, with the Entomology community and the disciplines represented at the NYS Agricultural Experiment Station at Geneva serving as pilot implementation subject areas. The task force should undertake the pilot in the context of also enhancing access to specialized subject resources through the Mann Library web site. In addition to exploring enhanced subject access, the task force should outline the best approach for continuing to provide a web presence for the Lee Library at Geneva, using the same technical environment as the Mann Library web site.

Public Services Executive Committee (PSEC)
PSEC represents the needs of the library user communities to the Library Administration. Its responsibility is to coordinate public services activities within CUL, recommend policy, define goals and objectives related to public services activities, and foster responsive and innovative services which may be delivered in person or virtually over the web. The Public Services mandate includes areas, such as instruction, research consultations, circulation, outreach, and public computing. PSEC coordinates its activities and collaborates with staff in areas such as Collection Building, Information Technology, Library Operations, and Research and Assessment.

Get It! Team
Originally created as a cross-divisional team charged with simplifying and streamlining patron access to library resources, the primary aim of the newly constituted Get It! Team remains unchanged, though the environment in which the team operates and the specific issues on which the group focuses have evolved since the original team's inception. Specifically, the Get It! Team is newly charged to:

- Participate in the testing of new and upgraded library discovery systems (e.g. the federated search system chosen to replace WebFeat later this year);
- Participate in the tracking and troubleshooting of user request failures using the CUL Web log software and other tools;
• Working with CLO Acquisitions and other staff, participate in the evaluation and testing of patron-driven acquisitions models;
• Serve as a forum for the discussion of interlibrary loan (ILL) issues and the dissemination of information related to ILL at Cornell;
• Participate in the review of all document delivery issues and policies, including those developed under the aegis of the 2CUL initiative;
• Play a leadership role in evaluating, implementing, and training associated with new and upgraded ILL software - including, but not limited to Borrow Direct, ILLiad, and RAPID;
• Communicate with reference and other CUL staff regarding patron access protocols and ILL issues; liaise with the Access Services Committee, the WorldCat Local Implementation Team, and other groups, as appropriate;
• Keep abreast of developments regarding patron access protocols and ILL issues, including attendance at local and national meetings, as appropriate.

Cataloging department
The mission of Cataloging and Metadata Services is to facilitate the discovery and delivery of information resources in support of the research and teaching needs of the scholarly community by applying expertise in the description and organization of materials.

Delivery & Metadata Management Services
(http://lts.library.cornell.edu/lts/dm/index)
DMMS acquires, delivers, and provides licensed access to information resources in all formats to CUL users in support of the teaching, learning, and research needs of Cornell University's scholarly community. DMMS also maintains and enhances the quality and scope of the metadata associated with these information resources through its database management operations, which include the development and support of automated methods for maintaining and enhancing access to these resources.

Cornell Library IT
CUL-IT Infrastructure division manages the technology infrastructure that undergirds the Library’s provision of services over the web. Infrastructure components include server hardware and software, enterprise-level systems, contracted data management services, digital media production facilities, public and staff desktop computing, and administrative computing. IT Infrastructure staff members participate in grant writing, R&D projects, assessment of services, and the creation of a technological vision in support of the Library’s programs. IT Infrastructure staff work closely with their counterparts in DLIT’s Application Development division on programmatic teams in order to respond to the Library's changing needs and priorities. Both divisions consider lifecycle management and sustainability principles as DLIT collectively assesses, refines, and expands its program areas.

Digital Scholarship Services
(https://confluence.cornell.edu/display/culdss/About+Digital+Scholarship+Services)
The Digital Scholarship Services at Cornell University Library facilitates partnerships to promote and assess the role of information and communication technologies in enhancing
The center offers services to support digitization of visual and textual content as well as creation of indexing (metadata) information to facilitate the discovery and management of digital resources. The Center facilitates the identification of services and tools to help faculty and researchers find services, tools, and partners in support of their projects with digital components. Also included in the program are services in support of electronic publishing and online repositories for storing and disseminating various outputs of research and scholarly communication. The program also facilitates discourse to explore the potential, implications, and consequences of technology-enhanced research and learning environments such as digital publishing, preservation of digital information, new media fluency, and intellectual property rights.

Curated List of Library Resources Project
(https://confluence.cornell.edu/display/curwebdev/CuLLR):
The CuLLR (Curated List of Library Resources) team is charged to design and implement a process whereby print and electronic resources in the library catalog (in the form of metadata), and from other sources to be determined, are extracted in accordance with the subject areas for a specific library. These resources may then be annotated to identify the subject areas for which the resource is useful and other attributes as determined by the project team. The process must include the ability to update data extracted from the catalog and other sources without destroying the existing annotations. The CuLLR project will be coordinated with the Discovery and Access Task Force, particularly with Group 3, to insure that our efforts are in line with the long-term plans of Cornell University Library (CUL).

This CuLLR project has 2 phases. The first is to define the resource types (e.g. databases, books, journals, subject guides) to include, the appropriate data sources (e.g library catalog, database), and what attributes should be associated with each resource (e.g. subject area, scope). This phase must be completed by May 1, 2011, in order to be able to implement the program changes prior to the start of the fall semester. The second phase is the actual implementation of the CuLLR data. The CuLLR technical staff will work with the unit libraries to define how the CuLLR data is to be used. We will begin with the Engineering Library (which must have something implemented by August 1, 2011), and then the Physical Sciences Library, followed by the remaining virtual and then non-virtual libraries.
# Appendix C.2

## High risk digital projects

<table>
<thead>
<tr>
<th>Collection name</th>
<th>Date created</th>
<th>Content</th>
<th>Delivery platform</th>
<th>Total sessions 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>KMODDL</td>
<td>2004-2005</td>
<td>moving image, qtvr, monograph, photographs</td>
<td>home grown</td>
<td>399,290</td>
</tr>
<tr>
<td>Glopad</td>
<td>1998-2006</td>
<td>moving image, qtvr, monograph, photographs</td>
<td>home grown</td>
<td>185,149</td>
</tr>
<tr>
<td>RMC Exhibitions (multiple)</td>
<td>Ongoing</td>
<td>images/text</td>
<td>html</td>
<td>862,945</td>
</tr>
<tr>
<td>C-Theory</td>
<td>2002</td>
<td>interactive site</td>
<td>home grown</td>
<td>13,948</td>
</tr>
<tr>
<td>Fuertes</td>
<td>1998</td>
<td>images/text</td>
<td>html</td>
<td>3,983</td>
</tr>
<tr>
<td>Plant Pathology</td>
<td>Unknown</td>
<td>images</td>
<td>html</td>
<td>Unknown</td>
</tr>
<tr>
<td>Blaschka Image Gallery</td>
<td>2009</td>
<td>images</td>
<td>html</td>
<td>Unknown</td>
</tr>
<tr>
<td>Human Ecology</td>
<td>1999-2003</td>
<td>images</td>
<td>html</td>
<td>11,489</td>
</tr>
<tr>
<td>Rose Goldsen</td>
<td>2004</td>
<td>multiple</td>
<td>hybrid</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
Appendix D: Survey of Commercial and Open Source Solutions

CHARGE

To survey the functions provided by existing commercial and open source solutions for discovery layers and tools to improve the discovery to access process. Some relevant functions include (but not limited to) access mechanisms, techniques for discovery, potential for inter-connectivity, options for value added tags by library, etc.
Discovery Layers & Systems
Encore

Vendor web site: [http://encoreforlibraries.com/products](http://encoreforlibraries.com/products)

Summary: Encore is completely ILS-independent and provides an integrated discovery application that responds to searches with current scholarly and popular articles, local books, e-books, specialized digital collections, institutional repositories, and more.

Overview of features: See chart for a full list, but here are some more features:

Article discovery
- Provides access to real-time content from any provider
- Unlimited number of databases
- Databases can be organized into categories

Contextualization
- Context sensitive searching
- Start search in a single box using a topic or a known title
- Search display can be article or catalog focus
- Displays an expanded list of facets (ex. journal title)
- Facets are presented just like the native facets of the databases
- Encore becomes the native facet through the use of the native APIs
- Provides support for institution specific resources based on IPs
- Given users can access resources he/she is allowed to access
- Allows the system to know where the user is based on IPs
- Users can switch between authorized profiles
- Library profiles are determined by the library

Searching/Research Tools
- Allows for power-users (librarian, grad. student)
- Displays the last 3 recently searched and recently viewed
- Keyword and title specific searching
- Searches can be expanded to Boolean
- Limits can be applied (e.g. eBooks, books (print), ..)

Book Cart (the language of this text can be changed)
- Any title in search list can be added
- Remains persistent search to search
- User can make batch request
- User can select specific volume in multi-volume displays of items
- Ability to save to MyList (link to MyMillennium record)
- Toggle between brief and full view
Library Implementation Links:

- Yale Law School Library: [http://encore.law.yale.edu/iii/encore/home?lang=eng](http://encore.law.yale.edu/iii/encore/home?lang=eng) (Pearl Skin)
- University of Lincoln Nebraska: [http://iris.unl.edu/](http://iris.unl.edu/) (Custom Colors)
- Wright State University Libraries: [http://www.libraries.wright.edu/](http://www.libraries.wright.edu/) (Links under Logo)
- Westerville Public Library: [http://www.westervillelibrary.org/](http://www.westervillelibrary.org/) (Direct to Book Links)
- Scottsdale Public Library: [http://library.scottsdaleaz.gov/](http://library.scottsdaleaz.gov/) (Tagging, Ratings, and Reviews)
- University of Lincoln Nebraska: [http://encore.unl.edu/iii/encore/search/C](http://encore.unl.edu/iii/encore/search/C) (Yahoo Images)

Other Resources: A number of libraries have created YouTube Videos related to their Encore implementation.

- Long Beach Public: Intro to Encore: [http://www.youtube.com/watch?v=4fr0KYUe6lU](http://www.youtube.com/watch?v=4fr0KYUe6lU)
- LBP Social Catalog: Adding Tags, Reviews, Ratings: [http://www.youtube.com/watch?v=65M5WOkFG_I&NR=1](http://www.youtube.com/watch?v=65M5WOkFG_I&NR=1)
- Binghamton University: [http://www.youtube.com/watch?v=cw2VGh53FIY&feature=related](http://www.youtube.com/watch?v=cw2VGh53FIY&feature=related)
- Luther College: [http://www.youtube.com/watch?v=vRrUn0lR-Wo](http://www.youtube.com/watch?v=vRrUn0lR-Wo)
- Deerfield Public Library: [http://www.youtube.com/watch?v=QJt4Y2qSsMg](http://www.youtube.com/watch?v=QJt4Y2qSsMg)
- CLICnet: [http://www.youtube.com/watch?v=jcNvSqwL_Y0](http://www.youtube.com/watch?v=jcNvSqwL_Y0)
Discovery Layers & Systems

Blacklight

Vendor web site: http://projectblacklight.org/

Summary: Blacklight is a free and open source ruby-on-rails based discovery interface (a.k.a. "next-generation catalog") especially optimized for heterogeneous collections. You can use it as a library catalog, as a front end for a digital repository, or as a single-search interface to aggregate digital content that would otherwise be siloed. Blacklight uses solr, an enterprise-scale index for its search engine. Blacklight features faceted browsing, relevance based searching (with the ability to locally control the relevancy algorithms), bookmarkable items, permanent URLs for every item, user tagging of items.

Overview of features: Blacklight provides support for non-MARC records, so you can search and display materials, such as images, geographic data, and EADs. There are different views by domain or by discipline (e.g. Music, Earth Sciences). There are also specific displays and behaviors by content type. Blacklight has some immediate benefits that greatly improve usability, such as improved facet performance and better "did you mean?" suggestions.

Library Implementation Links: The University of Virginia Library and Stanford University Library are the primary development partners, but you can see a current list of installations on our Examples page.

Other Resources: None
Other notes: Blacklight is a discovery layer, but not discovery system.
Appendix D

Discovery Layers & Systems
Endeca


Summary: This is very different from the other products. It has no content. It is a search and guided navigation tool that customers load their data into. They do have library customers (listed below) but their primary customer base is business and government.

Key Benefits of Endeca's Information Access Platform include:

- Innovative user experiences: Endeca is the leading faceted search provider for websites that rely on their user experience to maximize task completion rates. Endeca maintains its own user experience research group, which continues to innovate new features, and to tailor solutions for specific client needs.
- Customized application development: Endeca solutions are built using APIs that allow extensive, fine-grained controls, while using open standards to integrate with other application components and data sources. In addition, they include web-based controls so line-of-business users can configure applications without relying on IT intervention.
- Advanced text mining: The metadata that fuels faceted search can be augmented with text mining, provided by Endeca's own semantic technologies as well as deep integrations with third-party partners.
- Enterprise-class infrastructure: The IAP is designed for unmatched data volume, traffic, security, languages, and reliability.

Overview of features: See chart for more details.

Library Implementation Links: Many libraries are using Endeca including University of London – Queen Mary, University of Toronto, University of Technology, Sydney, and many others. The best known examples and earliest adopters of Endeca are from Phoenix Public Library (http://www.phoenixpubliclibrary.org/default.jsp) and North Carolina State University Library (http://www.lib.ncsu.edu/). More recently, the John F. Kennedy Presidential Library implemented Endeca's Publishing Suite with Page Builder and the SEO Module to power the search functionality of the digital archives. You can read about this implementation at http://www.endeca.com/en/news-and-events/press-releases/2010/jfk-presidential-library-partners-with-endeca-on-historic-build-of-nation-first-online-presidential-archives.html

Other Resources: None
Other notes: None
Vendor web site: http://www.serialssolutions.com/summon/

Summary: The Summon™ web-scale discovery service enables a familiar web-searching experience of the full breadth of content found in library collections—from books and videos to e-resources such as articles. It goes beyond federated search, beyond next-generation catalogs to create an all-new service for libraries. Through one simple search to a single unified index, the Summon service provides instant access to the breadth of authoritative content that's the hallmark of great libraries.

Overview of features:

- A single search -- Provides a Google-like search experience, allowing researchers to use one search box to discover credible and reliable library content.
- Exceptionally relevant results -- Delivers search results in a relevancy-ranked list so the most relevant results appear at the top of the list.
- Complete and instant access -- Incorporates the comprehensive knowledge of your library's holdings so that researchers can access their results including full-text of electronic articles.
- Hosted for easy support -- The Summon™ service is provided as a hosted service, making it an easy application to support with very little impact on library staff.
- Mobile Accessibility -- Instant searching of a library's collection through the Summon™ web-scale discovery service from a mobile device
- Easily integrated -- The Summon™ web-scale discovery service provides an open API that allows integration with existing applications, such as learning management systems.
- Results refinement -- Easily navigate and narrow search result sets using multiple methods, such as filtering, faceting and sorting.
- Clear, immediate results -- Valuable information is conveniently displayed within the results list, allowing for quick appraisal. This data includes abstracts, item location, online full text availability, and more.
- Full-text filter -- The Summon™ service provides time-efficient research for the busy student by allowing results set to be limited to items immediately available in full text online.
- Citation formatting -- Quickly cut and paste citations in your preferred style.
- Bibliographic information export -- Easily export citations to bibliographic management software applications such as EndNote, RefWorks and ProCite.
- Scholarship filter -- Giving you the ability to limit results to "scholarly" publications including peer-reviewed material.
- "Did you mean?" suggestions -- The intuitive Summon™ service provides alternative guidance for misspellings or low yield queries.
- Content and Coverage: Content comes from 6,800+ publishers and 94,000+ journal and periodical titles, with over 500 million items indexed in the centralized index---and the list is constantly growing.
• Pricing depends on FTE and # of resources
• SUMMON now integrated with Ulrich's for determining "Scholarly Resource" status.
• API Documentation: http://api.summon.serialssolutions.com/help/api/
• Also see chart for more details.

Library Implementation Links:

Has been adopted by several libraries. Notable are those who have not depended on only SUMMON as their solution, but have created mashups of different solutions to provide customized user interfaces to meet their needs.

• NCSU: http://www.lib.ncsu.edu/
  o Winner of cutting edge award from ALA
  o Conducted a full library website redesign in 2010
  o Use a Best Best feature at the top
  o NCSU was not willing to use the one search box that SUMMONS or EBSCO provide
  o They decided to go with tabs and usability testing revealed 75% of users selected the appropriate tab, however they were not changing their tab after they did not get what they were looking for.
  o Used the tab "ALL" however it could be misleading that users can search all of NCSU materials, which is still not the case
  o Search Usage data revealed the following click share:
    • All tab: 8%
    • Articles: 43.8%
    • Books 36.4%
    • Library website: 2.4%
  o Quick search technical info: http://www.lib.ncsu.edu/dli/projects/quicksearch/
• University of Michigan (used Drupal to build SUMMON into it) and did it in 40 days
• Villanova - https://library.villanova.edu/Find/

Other Resources: None
Other notes: Andrew Nagy, developer of Vufind, is now working for Serial Solutions SUMMON.
Discovery Layers & Systems
EBSCO Discovery

Vendor web site: [http://www.ebscohost.com/discovery](http://www.ebscohost.com/discovery)

Summary: EBSCO Discovery Service claims to have the widest and deepest set of metadata which provides users with access to an institution’s entire collection via a single, customizable entry point - creating an experience that is comprehensive, fast, and familiar.

Overview of features:

- Single search box access to the entire collection
- Guided Style Advanced Search screen with drop-down list of fields in which to search for specific terms
- Expeditious Result List display
- Result List limiters and facets in the left-hand column help to consolidate user refinement actions
- To the right of the Result List, search/content expansion tools: Related Information, access to integrated search (EHIS) connectors, and widgets
- New checkbox design allows users to select multiple source types, clusters and content providers/databases
- "Breadbox" tracking enables easy removal of all user refinements, including limiters and expanders
- CustomLinks: article-level linking to link resolvers, publisher direct links, ILL and more
- Book jacket image catalog enhancement
- "Available in Library" collection limiter (fit to your collections/subscribed services)
- Simple and Advanced Search Screen options
- Various Limiters, including:
  - [View Screenshot: Limiters]
    - Full text (electronic & print)
    - Peer-reviewed
    - Date Slider
    - Catalog-only Limiter
- Facets (e.g., subjects, sources, authors, etc.)
  - [View Screenshot: Facets]
- Open Integration features, including:
  - Skinnable Interface (colors)
  - Branding Options
  - Tool Bar Customization including link wording and destination
  - Importable & Exportable Widgets
- Persistent Linking & Bookmarkable URLs
- Enhanced Relevancy Ranking
- Citation Formatting (including APA, MLA, etc.)
  - [View Screenshot: Citation]
• Bibliographic Export Feature
  (to popular bibliographic management tools)
  [View Screenshot: Export Bibliographic Record]
• Available on your Mobile Device
• Search Box Builder tool to make an EDS simple search box available on any web page
• Ability to limit searches to available full text (electronic & print)
• Full-featured user experience (EBSCOhost®)
• Full-text searching
• Guest/non-authentication access
• Much More...

Content: EDS Base Index + Custom Components
The EDS Base Index represents content from approximately 20,000 providers (and growing) in addition to metadata from another 70,000 book publishers. Although constantly growing, the EDS Base Index currently provides metadata for:

• Nearly 50,000 Magazines & Journals
• Nearly 6 million Books
• Nearly 20,000 Conference Preceedings
• 825,000 CDs & DVDs
• Hundreds of thousands of additional information sources from various source-types, including:
  o Biographies
  o Health Reports
  o Newspapers
  o And More

The Most Comprehensive Collection of Metadata
EDS enables users to search the widest and deepest collection of metadata and quickly access the content to which they have rights. Content components available with EBSCO Discovery Service include:

• Detailed metadata (e.g., author-supplied abstracts, keywords, subjects, etc.) from far more content providers and publishers than any other discovery service
• Robust metadata from the most comprehensive collection of journals & magazines
• Complete indexing from EBSCOhost databases to which an institution subscribes (e.g., Academic Search, Business Source, CINAHL, Historical Abstracts, etc.)
• Complete indexing from important non-EBSCOhost databases (to which a customer subscribes), including resources from Alexander Street Press, LexisNexis, NewsBank, Readex, and many others
• Complete OPAC loaded directly into EDS (and searched along with all other EDS content), includes:
  o Real Time availability checks
  o Daily Updates
Book jacket images, book records, entertainment records, annotations, family keys, subject headings, demand information, awards, review citations, etc., for hundreds of thousands of publications
[View Screenshot: Enhanced Book Data]
Institutional archives/repositories that can be directly loaded into EDS and searched as part of the overall experience

EDS customers can build upon the base index by adding metadata representing their own unique collections including their catalog (OPAC), databases, archival collections and more—allowing them to extend and enhance the solution to best fit their needs.

**Library Implementation Links:** University of North Carolina, Greensboro presented on integration of EBSCO Discovery. See [http://library.uncg.edu/](http://library.uncg.edu/)

- Implemented in July 2010
- They chose EBSCO because a large majority of their databases are EBSCO host and students are already familiar with it.
- Their most important takeaway was that "put the search box everywhere" -
- They chose to place their box in a small space on top of every page. So, users do not have to return to the home page each time. It is part of the banner on top of the page.
- They did not mash up solutions on the results page. Their results are purely provided by EBSCO.

**Other Resources:** None
**Other notes:** As of March 2010, there was more content added to their index and little in the way of new functionality since last look. EBSCO Discovery includes a federated search sidebar for an extra cost. EBSCO claims to have an API, but no examples were available of libraries using it in production.

**Discovery Layers & Systems**

**Deep Web Technologies**

**Vendor web site:** [http://www.deepwebtech.com/index.html](http://www.deepwebtech.com/index.html)
**Summary:** Baseema and Jesse met with the president of the company, Abe Lederman, at ALA annual.

- Employ engagement managers
- Customers are listed at [http://www.deepwebtech.com/customers/index.html](http://www.deepwebtech.com/customers/index.html)
- Example: [Science.gov](http://www.deepwebtech.com/customers/index.html)
- Clustering is available but not full scale faceting
- Incremental results, so user can begin working with some while others are being retrieved
- API is available, so customer can build the front end
- The search builder is fairly easy to use and anybody can build the search
- Not yet compatible with course management systems
- Very flexible company and will work with the customer to customize solutions
Overview of features: See chart.

Library Implementation Links:

- Science.gov – [http://www.science.gov](http://www.science.gov) – A single search goes across several sources of records. Results appear as a primary list, with several related sets of results to the right (Wikipedia, EurekAlert.
- Stanford uses Deep Web for xSearch (behind institutional login)

Other Resources: None

Other notes: None
Discovery Layers & Systems
PRIMO

Vendor web site: http://www.exlibrisgroup.com/category/PrimoOverview

Summary: PRIMO is a discovery system for local and remote resources, such as books, journal articles, and digital objects. It provides indexing and abstracts of millions of e-resources through the PRIMO Central Index through its participating publisher program. PRIMO is already mobile device capable and can be enabled.

Overview of features: See the chart for more details than listed below.
- Provides indexing and abstracting through PRIMO Central Index
- Capable of meta searching across library websites, institutional repositories, external web content and restricted library resources
- Supports variety of digital, audio-visual and print formats
- Does not currently support a people discovery system, need to investigate the capabilities of the API
- Integrates with CUL’s ILS, Voyager
- Open resolver SFX available for integration, if we want to replace Get it!
- Usage data reporting available
- Likely offers an API that can be integrated into course management systems and other Cornell University websites. We might want to renew the conversation with Cornell University website office to place that API on the search page.
- Social features such as commenting, incorporating outside content such as Amazon reviews and sharing features are available
- Offers Bx Recommender, a tool specifically built for suggestions based on meta usage statistics
- Offers persistent URLs
- Offers faceted results
- Browsing sub-domain of holdings is possible
- APIS are available, according to vendor
- OAI toolkit available
- Mobile friendly interface available
- Look and feel: skinning to customize look available

Representative Library Implementation Links:

Vanderbilt Discover Library: http://discoverlibrary.vanderbilt.edu/
Royal Library of Denmark: http://www.kb.dk/en/
Discovery Layers & Systems
VuFind

Vendor web site: http://vufind.org/index.php

Summary: VuFind is a discovery layer that was originally created by Andrew Nagy, who went to work for Serial Solutions. It is open source and downloadable from the website. It is meant to replace the multiple discovery systems adopted by institutions making it possible to search catalog records, digital library items, Institutional Repository, and other library collections and resources. Its social features are fairly strong, however Worldcat has much more stronger social features than Vufind.

Overview of features:

- It is only a discovery layer and does not provide any indexing or abstracting content
- Capable of meta searching across library websites, institutional repositories, external web content and restricted library resources
- Supports variety of digital, audio-visual and print formats
- Does not currently support a people discovery system, need to investigate the capabilities of the API
- Records need to be exported out of Voyager and into VuFind
- It is completely modular and available through GPL open source license
- It purports to be flexible and customizable
- Strong social features amongst many products including ability for users to comment and bring in reviews from external sources
- Compatible with Zotero
- Some contextualization features for authors, but VIVO would far exceed the capabilities of VuFind
- Offer features such as cite this (APA, MLA), add to favorites and make lists, and share by emailing
- Offers persistent URLs
- Offers some faceted results
- APIS are available, according to vendor
- OAI toolkit available
- Text this feature allows user to send to their mobile device
- Look and feel: skinning to customize look available
- Also, see chart.

Representative Library Implementation Links: Villanova University Library: https://library.villanova.edu/

Discovery Layers & Systems
eXtensible Catalog

Vendor web site: http://www.extensiblecatalog.org/
Summary: eXtensible catalog is a discovery system being developed by University of Rochester. It intends to be an open source discovery system comprised of four software components that allow integration of library metadata, ILS circulation services, institutional repositories, and website content to be presented in a single interface. The metadata toolkit normalizes data to achieve consistency.

Overview of Features:

- Provides a single user interface for searching across ILS, digital repository, and library web page content. Users can find everything from library hours to books and journals in one place.
- Implements innovative faceted browse features
- Offers tools to build web applications that tie into the library’s ILS.
- Integrates fully with an existing ILS and authenticates the patron (unlike Worldcat where a separate login is needed for Worldcat and Voyager)
- Built on Drupal
- A library of extensions available

Representative Library Implementation Link:
University of Rochester:

Only demos are available through slide shows as of now. It seems like it is almost ready as seen in the ALA midwinter presentation slides available.
<table>
<thead>
<tr>
<th>Discovery Layer</th>
<th>Summon</th>
<th>Deep Web Technologies</th>
<th>Encore</th>
<th>EBSCO Discovery</th>
<th>Endeca</th>
<th>PRIMO</th>
<th>VuFind</th>
<th>eXtensible Catalog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indexing and abstracting services [meta]</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes. Article searches are integrated in real-time.</td>
<td>Yes</td>
<td>No</td>
<td>Yes (PRIMO Central Index)</td>
<td>No (it is a discovery layer)</td>
<td>No</td>
</tr>
<tr>
<td>Subject-specific research guides and webliographies [meta]</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Yes?</td>
<td>Yes, Widgets integrate nicely with products like LibGuides. Separate EBSCOhost Profiles can be configured by subject area.</td>
<td>No</td>
<td>Yes</td>
<td>Yes (<a href="http://discovery.library.colostate.edu/">http://discovery.library.colostate.edu/</a>)</td>
<td>Yes</td>
</tr>
<tr>
<td>Encyclopedia articles and dictionary entries [meta]</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Yes, thru the article integration functionality</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Journal articles</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes - deep article integration</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>News reporting and newspaper articles</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes - deep article integration</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Articles, essays, or chapters (edited books)</td>
<td>Yes</td>
<td>Yes*</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Book-length monographs (one subject)</td>
<td>Yes</td>
<td>Yes*</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Images, including Maps</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Unknown</td>
<td>Unknown, but we can gather web usage statistics</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Statistics and Data</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Unknown, but we can gather web usage statistics</td>
<td>Maybe, only in some business databases</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Discovery Layer</td>
<td>Summon</td>
<td>Deep Web Technologies</td>
<td>Encore</td>
<td>EBSCO</td>
<td>Endeca</td>
<td>ExLibris Primo</td>
<td>VuFind</td>
<td>Extensible Catalog</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------</td>
<td>-----------------------</td>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>----------------</td>
<td>--------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dissertations and theses</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multimedia (audio, video)</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Musical scores</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Yes</td>
<td>Maybe, if in the catalog</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Special collections materials: Manuscripts, letters, diaries, etc.</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Laws, patents, and government documents</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Conference papers and presentations</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Working papers and gray literature</td>
<td>Yes</td>
<td>Yes*</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>People: Subject specialists; reference and research staff</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Unknown</td>
<td>Maybe, if via Widget</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>Institutional Digital Repositories</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Functionality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integrated Library System (Voyager) - placing renewals, recalls</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes (also ILS-independent)</td>
<td>Yes</td>
<td>Unknown</td>
<td>Integrates with Voyager</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Open URL resolver</td>
<td>Unknown</td>
<td>Ties into local resolver</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes (SFX)</td>
<td>can use any other</td>
<td>can use any other</td>
</tr>
<tr>
<td>Usage Statistics reporting</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes - Google Analytics</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
</tr>
<tr>
<td>Meta search Functionality (other webpages)</td>
<td>Unknown</td>
<td>Yes*</td>
<td>Yes - can pass searches to WorldCat</td>
<td>Yes</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Yes</td>
</tr>
<tr>
<td>Integration with Course Management Systems</td>
<td>Unknown</td>
<td>No</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Discovery Layer</td>
<td>Summon</td>
<td>Deep Web Technologies</td>
<td>Encore</td>
<td>EBSCO</td>
<td>Endeca</td>
<td>ExLibris Primo</td>
<td>VuFind</td>
<td>Extensible Catalog</td>
</tr>
<tr>
<td>-----------------------------------------</td>
<td>--------------</td>
<td>-----------------------</td>
<td>---------</td>
<td>-----------</td>
<td>----------</td>
<td>----------------</td>
<td>--------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Functionality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social features (commenting,</td>
<td>No</td>
<td>Yes, but limited</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>incorporating outside content into the</td>
<td></td>
<td>features)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>catalog, e.g., Amazon reviews,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sharing features)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Faceted results</td>
<td>Yes</td>
<td>“Intelligent Clusters”,</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>but not fully faceted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suggestions (More like this)</td>
<td>Yes</td>
<td>Has alerts</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Browsing sub-domains of holdings</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persistent URLs</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On the Spot Translation</td>
<td>No</td>
<td>No</td>
<td>Does not translate but can labels in another language</td>
<td>Yes</td>
<td>Unknown</td>
<td>No</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APIs</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile friendly (send records to their</td>
<td>Yes</td>
<td>Unknown</td>
<td>Not at the moment</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes - PRIMO available for mobile phones</td>
<td>Yes - Text this feature for every catalog record</td>
<td>N/A</td>
</tr>
<tr>
<td>phone)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Source</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The OAI Toolkit (Harvesting ILS data)</td>
<td>Yes</td>
<td>No, but can federate</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>existing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Documentation</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Unknown</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metadata management platform</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Unknown</td>
<td>Solr Energy, Open Source Search Engine platform</td>
<td>Drupal</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>bx recommender</td>
</tr>
</tbody>
</table>

*As a federated search tool, Explorit should, in theory, be able to search any web-based system.*
Discovery Tools

Bx Recommender

- **Vendor web site:** [http://www.exlibrisgroup.com/category/bXOverview](http://www.exlibrisgroup.com/category/bXOverview)

- **Summary:** The tool provides scholarly recommendation service. Based on usage data and not keywords, it suggests highly granular recommendations to scholarly articles to our users through the SFX link resolver. All recommended articles are focused solely on the scholarly domain. In general, our patrons would receive current recommendations that are relevant: based on the analysis of millions of transactions related to the context of the user’s work and tightly integrated into scholarly search results. It gives the recommendation of a current scholarly material down to the article level. In the future, the tool plans to identify interdisciplinary connections that can inform selectors and identify scholarly trends. Much of bX’s functionality and design is based on research done as part of the MESUR project at Los Alamos National Laboratory by Johan Bollen and Herbert Van de Sompel.

- **Overview of features:**
The scholarly recommendation tool give the user:
  - Added value: Users are guided to information they need and offered the opportunity to broaden the horizons of their research.
  - Fast delivery: bX recommendations are delivered to Cornell’s system, so users can start receiving recommendations immediately.
  - Intuitive presentation: Cornell users view recommendations in their familiar search environment, depending on where we would implement the feature. It is our choice where to have user have the recommendations appear.
  - Based on a massive data set: Usage data harvested from link resolver logs and builds a very large aggregate of usage data. Currently with over 250 partners, there are 26 million articles in the data set and it is growing at 5 million per month. Also, the more institutions that contribute, the larger the usage data and stronger the recommendations
  - API available: Request to bx can be sent via API
  - Multiple results available: Results are returned as xml, text, ATOM, RSS, HTML

- **Library Implementation Links:** It is not a service that can be seen by non-allifiated library users, but libraries have implemented Bx in different ways such as:
  - University of Huddersfield Library has implemented Bx along with Summons.
  - Appears in the Find it box at Washington University St. Louis
  - Appears in Get it! midway in box at Vanderbilt Library
  - Sample view of the Bx recommender service
bX Recommendations

Users interested in this article also expressed an interest in the following:

4. SABOT, C Espaces de Dirichlet relia per des points et application aux diffusions sur les fractals finement ramifies Potential analysis 11 (1999) 103

View More...

- Other Resources:
  - There is a case study from the University of Plymouth and Ex Libris at http://www.exlibrisgroup.com/files/CaseStudy/PlymouthbXCaseStudy.pdf

- Other notes:
  - The cost is $3000 for Cornell.
  - There are some concerns about the privacy aspects of bX and with data harvesting, including the addition of monographs and other materials.
  - California State University, a subscriber to the tool, wishes bX had more recommendations (i.e., more institutions) and further experimentation with saved record recommendations and alerts.
**Discovery Tools**
**Masterkey**

- **Vendor web site:** [http://www.indexdata.com/masterkey](http://www.indexdata.com/masterkey)

- **Summary:** MasterKey is a growing and evolving family of tools for building sophisticated information discovery solutions. They can be used individually, or they can be combined together using a shared service-based architecture. Some of the tools have been released by Index Data under OSS (open source software) licenses, whereas others, at this time, are not open source. The most interesting part of the products offered by Indexdata is Pazpar2, a metasearch product.

- **Overview of features:**
  These are the current members of the family:

  ![Diagram of MasterKey architecture](image)

  - Pazpar2
    - Pazpar2 is a high-performance, user interface-independent, data model-independent metasearching middleware webservice featuring merging, relevance ranking, record sorting, and faceted results. It will search SOLR/Z39.50/SRU/SRW targets in parallel, create an internal, merged result set, and make this available to webservice clients.
  - Connector Framework
    - This is a technology for providing access to remote systems that do not offer standard APIs or protocols. Specifically, it can automate the kinds of interactions that end-users might have with a web-based interface, and construct structured output from presentation-formatted displays.
  - Zebra
    - Zebra is a metadata indexing system. It will ingest MARC or XML and index records according to arbitrarily complex rules.
Service Proxy -- This is a service that sits in front of Pazpar2 and augments the Pazpar2 web service API. It is designed to be a general-purpose service platform into which we can add modules to support an increasing array of higher-order functions. In addition to the core functionality, we are looking at adding features such as tagging and reviews of records, broadcast searching based on OpenURLs, circulation status lookup, etc.

User Interface Framework -- This is a simple JavaScript-based framework for developing interfaces for Pazpar2. The framework instantiates each of the main components in a typical interface (result list, facets, record view, etc.) as a 'widget' that can be placed anywhere on a page and configured with respect to how things are presented (like records, lines in facet views, etc.)

Torus -- The Torus is both a software component and architecture for maintaining complex knowledge bases in a distributed, cross-institutional fashion. Its primary use is to manage lists of search targets, either as global resources that are made available to customers, or as smaller lists that represent the selection of an individual customer.

MKAdmin -- This is an administrative shell designed to work with the Torus. It's both a practical example of an interface to the Torus platform, and a useful tool in its own right. It allows for the management of global search targets, local 'library' accounts with their own subsets of targets, as well as access control for local administrators to the admin interface.

Harvester -- The harvester is a general-purpose scheduler for ingesting remote data sources. At present it supports OAI-PMH, MARC bulk download, and simple harvesting of WWW jump-pages (e.g. A-Z lists).

Metaproxy -- The Metaproxy is a general-purpose information retrieval switchboard, diagnostic tool, load balancer, traffic regulator, performance enhancement tool, etc.

- **Library Implementation Links**: None
- **Other Resources**: Presentation called “Metasearching and local indexes: making it Just Work when two worlds collide” by MasterKey
  

- **Other notes**: None
Discovery Tools
PubGet

- **Web site:** [http://pubget.org/](http://pubget.org/)

- **Summary:** Pubget has two main functions of interest:

  A search engine for life science PDFs. Pubget makes scientific research easier by simplifying the process of finding, managing and analyzing scientific papers. The goal is to be a single click from a result to a pdf. The core solution at www.pubget.com provides article-level tools making content discovery, access and copyright management much easier for the user. Pubget’s corporate services offer relevant banner and contextual advertisement for marketers and search services for libraries and R&D departments, including repository and text mining platforms.

  A pilot project with OCLC WorldCat to improve display and utility of institution’s e-journal holdings in the WorldCat system. Cornell is currently involved in this, with Adam Chandler coordinating. It involves entering admin logon information for our e-journal provider administrative interfaces. PubGet system then logs in, finds the list of holdings. Periodic updates of holdings list provided to WorldCat. If successful, this will serve to provide users with accurate holdings information as per cied by our e-journal providers, therefore increasing the likelyhood of successful connection to the subscribed content.

- **Overview of features: Pilot Project:**
  - Interface for adding/editing admin info
  - Summary and detailed views of e-journal holdings
  - Usage statistics for e-journals (fee based service)
● **Library Implementation Links:** Here is a screenshot of the interface for Cornell:

![Screenshot of a library interface](image)

- **Upper-level Ontologies for Health Information Systems. Towards an Archetype Patterns Approach.**

- **Analysis of Expression Sequences Tags from the Plants of the Live-Releasing Path-Practiced Forests.**

- **A common layer of interoperability for biomedical ontologies based on OWL EL.**

- **Other Resources:** none

- **Other notes:** It is early in the Cornell pilot process, so results of the pilot are unclear. Adam Chandler is the lead. The goal of this is to improve our e-serials holdings data in WorldCat.
Discovery Tools
Umlaut


- **Summary:** Umlaut is OpenURL link resolving middleware that adds functions and services to commercial link resolving software such as SFX. The services also is capable of aggregating third party "know item services" and sharing them with other services such as the library catalog or federated search tool. Unlike most typical link resolver products (such as SFX), the Umlaut does not manage its own “knowledge base” of information on what titles an institution possesses from what vendors, and how to link to them. Umlaut relies on SFX--accessed through the SFX API--for that information and service. Umlaut is an open source project originally developed by Ross Singer while at Georgia Tech, and subsequently worked on quite a bit by Jonathan Rochkind of Johns Hopkins University.

- **Overview of features:**
  - Provides links to look up a periodical citation in Ulrich’s
  - Umlaut uses WorldCat Identities by extract metadata from the OpenURL or other services and sending it to WorldCat Identities to find more items by the author of the item being viewed.
  - Provides a great deal of flexibility with the user interface
  - Provides a ‘full’ link resolver menu too, which the user can choose to see, and which is shown when no full text is available
  - Adds functions and services to commercial link resolving software

- **Library Implementation Links:** Johns Hopkins University Library has Umlaut implemented. See [http://findit.library.jhu.edu/go/329602?umlaut.skip_resolve_menu=false](http://findit.library.jhu.edu/go/329602?umlaut.skip_resolve_menu=false) for an example.

- **Other Resources:** For more technical information, go to [http://wiki.code4lib.org/index.php/Umlaut_full_API](http://wiki.code4lib.org/index.php/Umlaut_full_API)

- **Other notes:** None
Discovery Tools
Carrot 2 Results clustering

- **Web site:** [http://project.carrot2.org/download.html](http://project.carrot2.org/download.html)

- **Summary:**
  Results clustering (Rick) - Carrot2 is an Open Source Search Results Clustering Engine. It can automatically organize small collections of documents, e.g. search results, into thematic categories. Carrot2 comes with java libraries, php libraries, and a commandline interface. There is a plugin for the browser. (Firefox, and Internet Explorer). There is a web application which uses the clustering server to create visualizations with flash or a 'conventional tree view'.

- **Overview of features:**
  - The tools has two specialized document clustering algorithms,
  - Offers ready-to-use components for fetching search results from various sources including YahooAPI, GoogleAPI, Bing API, eTools Meta Search, Lucene, SOLR, Google Desktop and more.
  - Enables modifying clustering algorithm’s attributes and observing the results in real time
  - Carrot2 can cluster content in 19 languages
  - It cannot cluster documents to some predefined clusters / labels

- **Library Implementation Links:** Unknown

- **Other Resources:** None

- **Other notes:** None
Discovery Tools
Boopsie

- **Web site:** [http://www.boopsie2.com/libraries.html](http://www.boopsie2.com/libraries.html)

- **Summary:** Develop mobile apps for various industries, including libraries. Boopsie is compatible with all web-enabled phones. High-speed downloadable clients are available for Android, BlackBerry, J2ME, Palm OS, Symbian S60, Windows Mobile and iPhone.

- **Overview of features:** Library app features include: Real-time ILS integration for patron account info, requests, etc...; MobileFind – enables fast catalog search with minimal keystrokes; GPS-aware Library Locator; Ask a Librarian; Reading Lists; Calendar & Events; Integrated social networking tools.

- **Library Implementation Links:** Current clients include Brown, Indiana University, Vanderbilt and others. Academic customers at: [http://www.boopsie2.com/libraries_clients.html#academic](http://www.boopsie2.com/libraries_clients.html#academic)

- **Other Resources:** None
- **Other notes:** None
APIs Available

Standard APIs

- SRU/W http://www.loc.gov/standards/sru/
- NCIP http://www.niso.org/workrooms/ncip
- SPARQL http://www.w3.org/TR/rdf-sparql-query/

Publisher Proprietary APIs

- Springer API http://dev.springer.com/docs
- Scopus API http://searchapidocs.scopus.com/
- Elsevier http://www.applications.sciverse.com/action/userhome
- Ebrary http://www.ebrary.com/corp/techAPI.jsp
- Wiley (this may be some sort of legacy api) http://www.wiley.com/legacy/authors/apifno/
- O'Reilly Safari http://access.safari.oreilly.com/affiliates/?p=web_services

Aggregator APIs

- Summon http://api.summon.serialssolutions.com/help/api/
- EBSCO Discovery, could not find on line docs. http://www.ebscohost.com/discovery
- Encore Synergy http://encoreforlibraries.com/products

Not Specific to any vendor (not a publisher).

- https://isbndb.com/docs/api/51-books.html
- CrossRef (DOI) API http://labs.crossref.org/site/quick_and_dirty_api_guide.html
- WorldCat API http://oclc.org/developer/documentation/worldcat-search-api/using-api
- LibraryThing http://www.librarything.com/services/
- Amazon http://aws.amazon.com/ They have lots of APIs, like Google.
- HathiTrust http://www.hathitrust.org/data

General Purpose APIs that might be useful

Others, less bibliographic

- ArtStor
- RefWorks

Peer Mashup Solutions

There are a few examples of libraries providing the results from several "silos" together on the initial results screen.

University of Michigan - http://www.lib.umich.edu/ - Drupal framework featuring results from Summon, local catalog, databases, Research guides, website, collections, library contacts, institutional repository, gov. docs, journals

NCSU - http://www.lib.ncsu.edu/ - ALL Tab gives results from Summon API, Local catalog, Library Website, Journal collections, and pointers to more results in the native systems or other searches. Their catalog search can eventually link to a WC-local instance for availability beyond NCSU.

Villanova - https://library.villanova.edu/Find/ - Side by side results from Summon and local catalog (I believe Voyager is underneath this). One click provides access to facets and other tools for the individual system. VU-Find is the display layer.