Investigations into Library Web-Scale Discovery Services

Jason Vaughan

ABSTRACT

Web-scale discovery services for libraries provide deep discovery to a library's local and licensed content and represent an evolution—perhaps a revolution—for end-user information discovery as pertains to library collections. This article frames the topic of web-scale discovery and begins by illuminating web-scale discovery from an academic library's perspective—that is, the internal perspective seeking widespread staff participation in the discovery conversation. This included the creation of the Discovery Task Force, a group that educated library staff, conducted internal staff surveys, and gathered observations from early adopters. The article next addresses the substantial research conducted with library vendors that have developed these services. Such work included drafting of multiple comprehensive question lists distributed to the vendors, onsite vendor visits, and continual tracking of service enhancements. Together, feedback gained from library staff, insights arrived at by the Discovery Task Force, and information gathered from vendors collectively informed the recommendation of a service for the UNLV Libraries.

INTRODUCTION

Web-scale discovery services, combining vast repositories of content with accessible, intuitive interfaces, hold the potential to greatly facilitate the research process. While the technologies underlying such services are not new, commercial vendors releasing such services, and their work and agreements with publishers and aggregators to preindex content, is very new. This article in particular frames the topic of web-scale discovery and helps illuminate some of the concerns and commendations related to web-scale discovery from one library's staff perspective—that is, the internal perspective. The second part focuses on detailed dialog with the commercial vendors, enabling the library to gain a better understanding of these services. In this sense, the second half is focused externally. Given that web-scale discovery is new for the library environment, the author was unable to find any substantive published work detailing identification, research, evaluation, and recommendation related to library web-scale discovery services. It's hoped that this article will serve as the ideal primer for other libraries exploring or contemplating exploration of these groundbreaking services.

Web-scale discovery services are able to index a variety of content, whether hosted locally or remotely. Such content can include library ILS records, digital collections, institutional repository content, and content from locally developed and hosted databases. Such capabilities existed, to varying degrees, in next-generation library catalogs that debuted in the mid 2000s. In addition, web-scale discovery services pre–index remotely hosted content, whether purchased or licensed by the library. This latter set of content—hundreds of millions of items—can include items such as e-books, publisher or aggregator content for tens of thousands of full-text journals, content from abstracting and indexing databases, and materials housed in open-access repositories. For purposes of this article, web-scale discovery services are flexible services which

Jason Vaughan (jason.vaughan@unlv.edu) is Director, Library Technologies, University of Nevada, Las Vegas.

provide quick and seamless discovery, delivery, and relevancy-ranking capabilities across a huge repository of content. Commercial web-scale discovery vendors have brokered agreements with content providers (publishers and aggregators), allowing them to pre–index item metadata and full-text content (unlike the traditional federated search model). This approach lends itself to extremely rapid search and return of results ranked by relevancy, which can then be sorted in various ways according to the researcher's whim (publication date, item type, full text only, etc.). By default, an intuitive, simple, Google-like search box is provided (along with advanced search capabilities for those wishing this approach). The interface includes design cues expected by today's researchers (such as faceted browsing) and, for libraries wishing to extend and customize the service, embraces an open architecture in comparison to traditional ILS systems.

Why Web-scale Discovery?

As illustrated by research dating back primarily to the 1990s, library discovery systems within the networked online environment have evolved, yet continue to struggle to serve users. As a result, the library (or systems supported and maintained by the library) is often not the first stop for research—or worse, not a stop at all. Users accustomed to a quick, easy, "must have it now" environment have defected, and research continues to illustrate this fact. Rather than weave these research findings into a paragraph or page, below are some illustrative quotes to convey this challenge. The quotations below were chosen because they succinctly capture findings from research involving dozens, hundreds, and in some cases thousands of participants or respondents:

People do not just use information that is easy to find; they even use information that they know to be of poor quality and less reliable—so long as it requires little effort to find—rather than using information they know to be of high quality and reliable, though harder to find.[1]

* * *

Today, there are numerous alternative avenues for discovery, and libraries are challenged to determine what role they should appropriately play. Basic scholarly information use practices have shifted rapidly in recent years, and as a result the academic library is increasingly being disintermediated from the discovery process, risking irrelevance in one of its core functional areas [that of the library serving as a starting point or gateway for locating research information] . . . we have seen faculty members steadily shifting towards reliance on network-level electronic resources, and a corresponding decline in interest in using locally provided tools for discovery.[2]

* * *

A seamless, easy flow from discovery through delivery is critical to end users. This point may seem obvious, but it is important to remember that for many end users, without the delivery of something he or she wants or needs, discovery alone is a waste of time.[3]

* * *

End users' expectations of data quality arise largely from their experiences of how information is organized on popular Web sites. . .[4]

* * *

[User] expectations are increasingly driven by their experiences with search engines like Google and online bookstores like Amazon. When end users conduct a search in a library

catalog, they expect their searches to find materials on exactly what they are looking for; they want relevant results.[5]

<center>* * *</center>

Users don't understand the difference in scope between the catalog and A&I services (or the catalog, databases, digitized collections, and free scholarly content).[6]

<center>* * *</center>

It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of "silo" systems whose distinctions seem arbitrary . . . the continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings. Our users expect simplicity and immediate reward and Amazon, Google, and iTunes are the standards against which we are judged. Our current systems pale beside them.[7]

<center>* * *</center>

Q: If you could provide one piece of advice to your library, what would it be?

A: Just remember that students are less informed about the resources of the library than ever before because they are competing heavily with the Internet.[8]

Additional factors sell the idea of web-scale discovery. Obviously, something must be discoverable for it to be used (and of value) to a researcher; ideally, content should be easily discoverable. Since these new services index content that previously was housed in dozens or hundreds of individual silos, they can greatly facilitate the search process for many research purposes. Libraries often spend large sums of money to license and purchase content, sums that often increase annually. Any tool that holds the potential to significantly increase the discovery and use of such content should cause libraries to take notice. At time of writing, early research is beginning to indicate that these tools can increase discovery. Doug Way compared link-resolver-database and full-text statistics prior to and after Grand Valley State University's implementation of the Summon web-scale discovery service.[9] His research suggests that the service was both broadly adopted by the University's community and that it has led to an increase in their library's electronic resource discovery and use. Willamette University implemented WorldCat Local, and Bill Kelm presented results that showed an increase in both ILL requests as well as use of the library's electronic resources.[10] From another angle, information-literacy efforts focus on connecting users to "legitimate" content and providing researchers the skills to identify content quality and legitimacy. Given that these web-scale discovery services include or even primarily focus on indexing a large amount of scholarly research, such services can serve as another tool in the library's arsenal. Results retrieved from these services—largely content licensed or purchased by libraries—is accurate, relevant, and vetted, compared to the questionable or opinionated content that may often be returned through a web search engine query. Several of the services currently allow a user to refine results to just categorized as peer-reviewed or scholarly.

The Internal Academic Library Perspective: Genesis of the UNLV Libraries Discovery Task Force

The following sections of this article begin with a focus on the internal UNLV Library perspective—from early discussions focused on the broad topic of discovery to establishing a task

force charged to identify, research, evaluate, and recommend a potential service for purchase. Throughout this process, and as detailed below, communication with and feedback from the variety of library staff was essential in ensuring success. Given the increasing vitality of content in electronic format, and the fact that such content was increasingly spread across multiple access points or discovery systems, in late 2008 the University of Nevada Las Vegas (UNLV) Libraries began an effort to engage library staff in information discovery and how such discovery would ideally occur in the future. Related to the exponential growth of content in electronic format, traditional technical-services functions of cataloging and acquisitions were changing or would soon change, not just at UNLV, but throughout the academic library community. Coinciding with this, the Libraries were working on drafting their 2009–11 strategic plan and wanted to have a section highlighting the importance of information discovery and delivery with action items focused on improving this critical responsibility of libraries. In spring 2009, library staff were given the opportunity to share with colleagues a product or idea, related to some aspect of discovery, which they felt was worthy of further consideration. This event, open to UNLV Libraries staff and other Nevada colleagues, was titled the Discovery Mini-Summit, and more than a dozen participants shared their ideas, most in a poster-session format.

One of the posters focused on Serial Solutions Summon, an early entrant into the vendor web-scale discovery service landscape. At the time, it was a few months from public release. Other posters included topics such as the Flickr Commons (cultural heritage and academic institutions exposing their digital collections through this popular platform), and a working prototype of a homegrown, open-source federated search approach searching across various subscribed databases. In August 2009, the dean of the UNLV University Libraries charged a ten-person task force to investigate and evaluate web-scale discovery services with the ultimate goal of providing a final recommendation for potential purchase. Representation on the task force included three directors and a broad cross section of staff from across the functional areas of the library, including back-of-the-house and public-service operations. The director of Library Technologies, and author of this article, was tasked with drafting a charge and chairing the committee; once charged, the Discovery Task Force worked over the next fifteen months to research, evaluate, and ultimately provide a recommendation regarding a web-scale discovery service. To help illustrate some of the events described, a graphical timeline of activities is presented as appendix A; the original charge appears as appendix B.

In retrospect, the initial target date of early 2010 to make a recommendation was naive, as three of the five products ultimately identified and evaluated by the task force weren't publicly released until 2010. Several boundaries were provided within the charge, including the fact that the task force was not investigating and evaluating traditional federated search products. The Libraries had had a very poor experience with federated search a few years earlier, and the shortcomings of the traditional federated search approach—regardless of vendor—are well known. The remainder of this article discusses the various steps taken by the Discovery Task Force in evaluating and researching web-scale discovery services. While many libraries have begun to implement the web-scale discovery services evaluated by this group, many more are currently at the learning and evaluation stage, or have not yet begun. Many libraries that have already implemented a commercial service likely went through an evaluation process, but perhaps not at the scale conducted by the UNLV Libraries, if for no other reason than the majority of commercial services are extremely new. Even in early 2010, there was less competition, fewer services to evaluate,

fewer vendors to contact, and fewer early adopters from whom to seek references. Fortunately, the initial target date of early 2010 for a recommendation was a soft target, and the Discovery Task Force was given ample time to evaluate the products. Based on presentations given by the author in 2010, it can't be presumed that an understanding of web-scale discovery—or the awareness of the commercial services now available—is necessarily widespread. In that sense, it's the author's hope and intent that information contained in this article can serve as a primer, or a recipe, for those libraries wishing to learn more about web-scale discovery and perhaps begin an evaluation process of their own.

While research exists on federated search technologies within the library environment, the author was unable to find any peer-reviewed published research on the evaluation model and investigations for vendor produced web-scale discovery services as described in this paper. However, some reports are available on the open web, providing some insights into web-scale discovery evaluations led by other libraries, such as two reports provided by Oregon State University. The first, dated March 2009, describes a task force whose activities included "scrutinize WCL [WorldCat Local], investigate other vendors' products, specifically Serials Solutions' Summon, the recently announced federated index discovery system; EBSCO's Integrated Search; and Innovative Interfaces' Encore product, so that a more detailed comparison can be done," and "by March 2010, communicate . . . whether WCL or another discovery service is the optimal purchase for OSU Libraries."[11] Note that in 2009, Encore existed as a next-generation discovery layer, and it had an optional add on called "Encore Harvester," which allows for the harvesting of digital local collections. The report cites the University of Michigan's evaluation of WCL, and adds their additional observations. The March 2009 report provides a features comparison matrix for WorldCat Local, Encore, Summon, and LibraryFind (an open-source search tool developed at OSU that provides federated searching for selected resources). Feature sets include the areas of search and retrieval, content, and added features (e.g., book covers, user tagging, etc.). The report also describes some usability testing involving WCL and integration with other local library services. A second set of investigations followed "in order to provide the task force with an opportunity to more thoroughly investigate other products" and is described in a second report provided at the end of 2009.[12] At the time of both phases of this evaluation (and drafted reports) three of the web-scale discovery products had yet to enter public release. The December 2009 report focused on the two released products, Serials Solutions Summon and WorldCat Local, and includes a feature matrix like the earlier report, with the added feature set of "other," which included the features of "clarity of display," "icons/images," and "speed." The latter report briefly describes how they obtained subject librarian feedback and the pros and cons observed by the librarians in looking at Summon. It also mentions obtaining feedback from two early adopters of the Summon product, as well as obtaining feedback from librarians whose library had implemented WorldCat Local. Apart from the Oregon reports, some other reports on evaluations (or selection) of a particular service, or a set of particular services, are available, such as the University of Michigan's Article Discovery Working Group, which submitted a final report in January 2010.[13]

Activity: Understanding Web-scale

The first activity of the Discovery Task Force was to educate the members, and later, other library colleagues, on web-scale discovery. Terms such as "federated search," "metasearch," "next

generation catalogs," and "discovery layers" had all come before, and "web-scale" was a rather new concept that wasn't widely understood. The Discovery Mini Summit served as a springboard that perhaps more by chance than design introduced to UNLV Library staff what would later become more commonly known as web-scale discovery, though even we weren't familiar with the term back in Spring 2009. In Fall 2009, the Discovery Task Force identified reports from entities such as OCLC, Ithaka, and reports prepared for the Library of Congress highlighting changing user behavior and expectations; these reports helped form a solid foundation for understanding the "whys" related to web-scale discovery. Additional registration and participation in sponsored web-scale discovery webcasts and meeting with vendors at library conferences helped further the understanding of web-scale discovery. After the Discovery Task Force had a firm understanding of web-scale discovery, the group hosted a forum for all library staff to help explain the concept of web-scale discovery and the role of the Discovery Task Force. Specifically, this first forum outlined some key components of a web-scale discovery service, discussed research the task force had completed to date, and outlined some future research and evaluation steps. A summary of these steps appears in the timeline in appendix A. Time was allowed for questions and answers, and then the task force broadcast several minutes of a (then recent) webcast talking about web-scale discovery.

As part of its education role, the Discovery Task Force set up an internal wiki-based webpage in August 2009 upon formation of the group, regularly added content, and notified staff when new content was added. A goal of the task force was to keep the evaluative process transparent, and over time the wiki became quite substantial. Links to "live" services were provided on the wiki. Given that some services had yet to be released, some links were to demo sites or sites of the closest approximation available, i.e., some services yet to be released were built on an existing discovery layer already in general release, and thus the look, feel, and functionality of such services was basically available for staff review. The wiki also provided links to published research and webcasts on Web-scale discovery. Such content grew over time as additional web-scale discovery products entered general release. In addition to materials on particular services, links were provided to important background documents and reports on topics related to the user discovery experience and user expectations for search, discovery, and delivery. Discovery Task Force meeting notes and staff survey results were posted to the wiki, as were evaluative materials such as information on the content-overlap analysis conducted for each service. Announcements to relevant vendor programs at the American Library Association's Annual Conference were also posted to the wiki.

Activity: Initial Staff Survey

As noted above, when the task force began its work, only two products (out of five ultimately evaluated) were in general release. As more products entered public release, a next step was to invite vendors onsite to show their publicly released product, or a working, developed prototype nearing initial public release. To capture a sense of the library staff ahead of these vendor visits, the Discovery Task Force conducted the first of two staff surveys. The 21-question survey consisted of a mix of "rank on a scale" questions, multiple-choice questions, and free-text response questions. Both the initial and subsequent surveys were administered through the online SurveyMonkey tool. Respondents were allowed to skip any question they wished. The survey was broken into three broad topical areas: "local library customization capabilities," "end user aspect:

features and functionality," and "content." The survey had an average response rate of 47 staff, or 47% of the library's 100-strong workforce. The survey questions appear in appendix C. In hindsight, some of the questions could have benefitted from more careful construction. That said, there was a conscious juxtaposition of differing concepts within the same question—the task force did not want to receive a set of responses in which all library staff felt it was important for a service to do everything—in short, to be all things to all people. Forcing staff to rate varied concepts within a question could provide insights into what they felt was really important. A brief summary of some key questions for each section follows. As an introduction, one question in the survey asked staff to rate the relative importance of each overarching aspect related to a discovery service (customization, end user interface, and content). Staff felt content was the most critical aspect of a discovery service, followed by the end-user interface, followed by the ability to heavily customize the service. A snapshot of some of the capabilities library staff thought were important (or not) is provided in table 1.

Web-scale Capabilities	SA	A	N	D	SD
Physical item status information	81.6%	18.4%	-	-	-
Publication date sort capability	75.5%	24.5%	-	-	-
Display library-specified links in the interface	69.4%	30.6%	-	-	-
One-click retrieval of full-text items	61.2%	36.7%	-	-	2%
Ability to place ILL / consortial catalog requests	59.2%	36.7%	4.1%	-	-
Display the library's logo	59.2%	36.7%	4.1%	-	-
To be embedded within various library website pages	58%	42%	-	-	-
Full-text items first sort capability	58.3%	31.3%	8.3%	2.1%	-
Shopping cart for batch printing, emailing, saving	55.1%	44.9%	-	-	-
Faceted searching	48.9%	42.6%	8.5%	-	-
Media type sort capability	47.9%	43.8%	4.2%	4.2%	-
Author name sort capability	41.7%	37.5%	18.8%	2.1%	-
Have a search algorithm that can be tweaked by library staff	38%	36%	20%	4%	2%
User account for saved searches and marked items	36.7%	44.9%	14.3%	4.1%	-
Book cover images	25%	39.6%	20.8%	10.4%	4.2%
Have a customizable color scheme	24%	58%	16%	2%	-
Google Books preview button for book items	18.4%	53.1%	24.5%	4.1%	-
Tag cloud	12.5%	52.1%	31.3%	4.2%	-
User authored ratings	6.4%	27.7%	44.7%	12.8%	8.5%
User authored reviews	6.3%	20.8%	50%	12.5%	10.4%
User authored tags	4.2%	33.3%	39.6%	10.4%	12.5%

SA = Strongly Agree; A = Agree; N = Neither Agree nor Disagree; D = Disagree; SD = Strongly Disagree

Table 1. Web-scale Discovery Service Capabilities

None of the results was surprising, other than perhaps the low interest or indifference in several Web 2.0 community features, such as the ability for users to provide ratings, reviews, or tags for items, and even a tag cloud. The UNLV Libraries already had a next-generation catalog offering these features, and they have not been heavily used. Even if there had been an appreciable adoption of these features by end users in the next-generation catalog for a web scale discovery service they are perhaps less applicable—it's probably more likely that users would be less inclined to post reviews and ratings for an article, as opposed to a monograph—and article-level content vastly outnumbers book-level content with web-scale discovery services.

The final survey section focused on content. One question asked about the incorporation of ten different information types (sources) and asked staff to rank how important it was that a service include such content. Results are provided in table 2. A bit surprisingly, inclusion of catalog records was seen as most important. Not surprisingly, full-text and A&I content from subscription resources were ranked very highly. It should also be noted that at the time of the survey, the institutional repository was in its infancy with only a few sample records, and awareness of this resource was low among library staff. Another question listed a dozen existing publishers (e.g., Springer, Elsevier, etc.) deemed important to the libraries and asked staff to rank the importance that a discovery service index items from these publishers on a four point scale from "essential" to "not important." Results showed that all publishers were ranked as essential and important. Related to content, 83.8 percent of staff felt that it was preferable for a service to de-dupe records such that the item appears once in the returned list of results; 14.6 percent preferred that the service not de-dupe results.

Information Source	Rating Average
ILS catalog records	1.69
Majority of full-text articles / other research contained in vendor-licensed online resources	2.54
Majority of citation records for non-full-text vendor-licensed A&I databases	4.95
Consortial catalog records	5.03
Electronic reserves records	5.44
Records within locally created and hosted databases	5.64
Digital collection records	5.77
WorldCat records	6.21
ILS authority control records	6.5
Institutional repository records	6.68

Table 2. Importance of Content Indexed in Discovery Service

After the first staff survey was concluded, the Discovery Task Force hosted another library forum to introduce and "test drive" the five vendor services in front of library staff. This session was scheduled just a few weeks ahead of the onsite vendor visits to help serve as a primer to engage library staff and get them actively thinking about questions to ask the vendors. The task force

distributed notecards at the forum and asked attendees to record any specific questions they had about a particular service. After the forum, specific questions related to the particular products were collected; 28 questions were collected, and they helped inform future research for those questions for which the task force did not at the time have an answer. Questions ran the gamut and collectively touched on all three areas of evaluation.

Activity: Second Staff Survey

Within a month after the five vendor onsite visits, a content analysis of the overlap between UNLV licensed content and content indexed by the discovery services was conducted. After these steps, a second staff survey was administered. This second staff survey had questions focused on the same three functional areas as the first staff survey: local library customization features, end user features and functionality, and content. Since the vendor visits had taken place and users could now understand the questions in the context of the products, questions were asked from the perspective of each product, e.g., "Please rate on a five point Likert scale whether each discovery service appears to adequately cover a majority of the critical publisher titles (WorldCat Local, Summon, EDS, Encore Synergy, Primo Central)." In addition, there were free-text questions focused on each individual product allowing colleagues to share additional, detailed thoughts. The second survey totalled 25 questions and had an average response rate of 18 respondents, or about 18 percent of library staff. Several staff conducted a series of sample searches in each of the services and provided feedback of their findings. Though this was a small response rate, two of the five products rose to the top, a third was a strong contender, and two were seen as less desirable. The lower response rate is perhaps indicative of several things. First, not all staff had attended the onsite vendor demonstrations or had taken the time to test drive the services via the links provided on the Discovery Task Force wiki site. Second, some questions were more appropriately answered by a subset of staff. For example, the content questions might best be matched to those with reference, collection development, or curriculum and program liaison duties. Finally, intricate details emerged once a thorough analysis of the vendor services was commenced. The first survey was focused more on the philosophy of what was desirable; the second survey took this a step further and asked how well each product matched such wishes. Discovery services are changing rapidly with respect to interface updates, customization options, and scope of content. As such, and also reflective of the lower response rate, the author is not providing response information nor analysis for this second survey within this article. However, results may be provided upon specific request to the author. The questions themselves for the second staff survey are significant, and they could help serve as a model for other libraries evaluating existing services on the market. As such, questions appear in appendix D.

Activity: Early Adopter References

One of the latter steps in the evaluation process from the internal academic library perspective was to obtain early adopter references from other academic library customers. A preliminary shortlist was compiled through a straw vote of the Discovery Task Force—and the results of the vote showed a consensus. This vote narrowed down the Discovery Task Force's list of services still in contention for a potential purchase. This shortlist was based on the growing mass of research conducted by the Discovery Task Force and informed by the staff surveys and feedback to date. Three live customers were identified for each service that had made the shortlist, and the task

force successfully obtained two references for each service. Reference requests were intensive and involved a set of two dozen questions that references either responded to in writing or answered during scheduled conference calls. To help libraries conducting or interested in conducting their own evaluation and analysis of these services, this list of questions appears in appendix E. The services are so new that the live references weren't able to comprehensively answer all the questions—they simply hadn't had sufficient time to fully assess the service they'd chosen to implement. Still, some important insights were gained about the specific products and, at the larger level, discovery services as a whole. As noted earlier, discovery services are changing rapidly in the sense of interface updates, customization options, and scope of content. As such, the author is not providing product specific response information or analysis of responses for each specific product—such investigations and interpretations are the job of each individual library seriously wishing to evaluate the services to help decide which product seems most appropriate for its particular environment. Several broad insights merit notice, and they are shared below.

Regarding a question on implementation (though some challenges were mentioned with a few responders), nothing reached the threshold of serious concern. All respondents indicated the new discovery service is already the default or primary search box on their website. One section of the early adopter questions focused on content. The questions in this area seemed a bit challenging for the respondents to provide lots of detail. In terms of "adequately covering a majority of the important library titles," respondents varied from "too early to tell," "it covers many areas but there are some big names missing," to two of the respondents answering simply, "yes." Several respondents also clearly indicated that the web-scale discovery service is not the "beginning and ending" for discovery, a fact that even some of the discovery vendors openly note. For example, one respondent indicated that web-scale discovery doesn't replace remote federated searching. A majority (not all) of the discovery vendors also have a federated search product that can, to varying degrees, be integrated with their preharvested, centralized, index-based discovery service. This allows additional content to be searched because such databases may include content not indexed within the web-scale discovery service. However, many are familiar with the limitations of federated search technologies: slow speed, poor relevancy ranking of results, and the need to configure and maintain sources and targets. Such problems remain with federated search products integrated with web-scale discovery services. Another respondent indicated they were targeting their discovery service at undergraduate research needs. Another responded, "As a general rule, I would say the discovery service does an excellent job covering all disciplines. If you start really in-depth research in a specific discipline, it starts to break down. General searches are great . . . dive deeper into any discipline and it falls apart. For example, for a computer science person, at some point they will want to go to ACM or IEEE directly for deep searches." Related to this, "the catalog is still important, if you want to do a very specific search for a book record, the catalog is better. The discovery service does not replace the catalog." In terms of satisfaction with content type (newspapers, articles, proceedings, etc.), respondents seemed generally happy with the content mix. A range of responses were received, such as "doesn't appear to be a leaning one way or another, it's a mix. Some of these things depend on how you set the system up, as there is quite a bit of flexibility; the library has to make a decision on what they want searched." Another example was that "the vendor has been working very hard to balance content types and I've seen a lot of improvement," "no imbalance, results seem pretty well rounded." Another responded, "A common complaint is that newspapers and book reviews dominate the search results, but that is much more a function of search algorithms then the amount of content in the index."

When asked about positive or critical faculty feedback to the service, several respondents indicated they hadn't had a lot of feedback yet. One indicated they had anecdotal feedback. Another indicated they'd received backlash from some users who were used to other search services (but also added that it was no greater than backlash from any other service they'd implemented in the past—and so the backlash wasn't a surprise). One indicated "not a lot of feedback from faculty, the tendency is to go to databases directly, librarians need to instruct them in the discovery service." For student feedback, one indicated, "We have received a few positive comments and see increased usage." Another indicated, "Reviews are mixed. We have had a lot of feedback thanking us for providing a search that covers articles and books. They like the ability to do one search and get a mix of resources without the search taking a long time. Other feedback usually centers around a bug or a feature not working as it should, or as they understand it should. In general, however, the feedback has been positive." Another replied, "Comments we receive are generally positive, but we've not collected them systematically." Some respondents indicated they had done some initial usability testing on the initial interface, but not the most recent one now in use. Others indicated they had not yet conducted usability testing, but it was planned for later in 2010 or 2011.

In terms of their fellow library staff and their initial satisfaction, one respondent indicated,

> "Somewhere between satisfied and very satisfied . . . it has been increasing with each interface upgrade . . . our instruction librarians are not planning to use the discovery service this fall [in instruction efforts] because they need more experience with it . . . they have been overall intrigued and impressed by it . . . I would say our organization is grappling more with the implications of a discovery tools as a phenomenon than with our particular discovery service in particular. There seems to be general agreement that it is a good search tool for the unmediated searcher."

Another indicated some concerns with the initial interface provided: "If librarians couldn't figure it out, users can't figure it out." Another responded, it was

> "a big struggle with librarians getting on board with the system and promoting the service to students. They continually compare it against the catalog. At one point, they weren't even teaching the discovery service in bib instruction. The only way to improve things it with librarian feedback; it's getting better, it has been hard. Librarians have a hard time replacing the catalog and changing things that they are used to."

In terms of local customization, responses varied; some libraries had done basically no customization to the out-of-the-box interface, others had done extensive customization. One indicated they had tweaked sort options and added widgets to the interface. Another indicated they had done extensive changes to the CSS. One indicated they had customized the colors, added a logo, tweaked the headers and footers, and created "canned" or preconfigured search boxes searching a subset of the index. Another indicated they couldn't customize the header and footer to the degree they would have liked, but were able to customize these elements to a degree. One respondent indicated they'd done a lot of customization to an earlier version of the interface, which had been rather painstaking, and that much of this broke when they upgraded to the latest version. That said, they also indicated the latest version was much better than the previous version. One respondent indicated it would be nice if the service could have multiple sources for

enriched record content so that better coverage could be achieved. One respondent indicated they were working on a complete custom interface from scratch, which would be partially populated with results from the discovery service index (as well as other data sources).

A few questions asked about relevancy as a search concept and how well the respondents felt about the quality of returned results for queries. One respondent indicated, "we have been able to tweak the ranking and are satisfied at this point." Another indicated, "overall, the relevance is good – and it has improved a lot." Another noted, "known item title searching has been a problem . . . the issues here are very predictable – one word titles are more likely to be a problem, as well as titles with stopwords," and noted the vendor was aware of the issue and was improving this. One noted, "we would like to be able to experiment with the discovery service more – and noted, "no relevancy algorithm control." Another indicated they looked to investigate relevance more once usability studies commenced, and noted they worked with the vendor to do some code changes with the default search mechanism. One noted that they'd like to be able to specify some additional fields that would be part of the algorithm associated with relevancy. Another optimistically noted "as an early adopter, it has been amazing to see how relevance has improved. It is not perfect, but it is constantly evolving and improving."

A final question asked simply, "Overall, do you feel your selection of this vendor's product was a good one? Do you sense that your users – students and faculty – have positively received the product?" For the majority of responses, there was general agreement from the early adopters that they felt they'd made the right choice. One noted that it was still early and the evaluation is still a work in progress, but felt it has been positively received. The majority were more certain, "yes, I strongly feel that this was the right decision . . . as more users find it, I believe we will receive additional positive feedback," "yes, we strongly believe in this product and feel it has been adopted and widely accepted by our users," "I do feel it was a good selection."

The External Perspective: Dialog with Web-scale Discovery Vendors

The preceding sections focused on an academic library's perspective on web-scale discovery services—the thoughts, opinions, preferences, and vetting activities involving library staff. The following sections focus on the extensive dialog and interaction with the vendors themselves, regardless of the internal library perspective, and highlight the thorough, meticulous research activities conducted on five vendor services. The Discovery Task Force sought to learn as much about the each service as possible, a challenging proposition given the fact that at the start of investigations, only two of five services had been released, and, unsurprisingly, very little research existed. As such, it was critical to work with vendors to best understand their services, and how their service compared to others in the marketplace. Broadly summarized efforts included identification of services, drafting of multiple comprehensive question lists distributed to the vendors, onsite vendor visits, and continual tracking of service enhancements.

Activity: Vendor Identification

Over the course of a year's work, the Discovery Task Force executed several steps to systematically understand the vendor marketplace—the capabilities, content considerations, development cycles, and future roadmaps associated with five vendor offerings. Given that the

task force began their work when only two of these services were in public release, there was no manual, recipe, or substantial published research to rely on. The beginning, for the UNLV Libraries, lie in identification of the services—one must first know the services to be evaluated before evaluation can commence. As mentioned previously, the Discovery Mini-Summit held at the UNLV Libraries highlighted one product—Serial Solutions Summon; the only released product at the time of the Mini-Summit was WorldCat Local. While no published peer-reviewed research highlighting these new web-scale discovery services existed, press and news releases did exist for the three to-be-released services. Such releases shed light on the landscape of services that the task force would review—a total of five services, from the first-to-market, WorldCat Local, to the most recent entrant, Primo Central.

OCLC WorldCat Local, released in November 2007, can be considered the first web-scale discovery service as defined in this research; the experience of an early pilot partner (the University of Washington) is profiled in a 2008 issue of *Library Technology Reports.*[14] In the UW pilot, approximately 30 million article-level items were included with the WorldCat database. Another product, Serials Solutions Summon, was released in July 2009, and together these two services were the only ones publicly released when the Discovery Task Force began its work. The task force identified three additional vendors each working on their own version of a web-scale discovery service; each of these services would enter initial general release as the task force continued its research: EBSCO EDS in January 2010, Innovative Interfaces Encore Synergy around May 2010, and Ex Libris Primo Central in June 2010. While each of these three were new in terms of web-scale discovery capabilities, each was built, at least in part, on earlier systems from the vendors. EDS draws heavily from the EBSCOhost interface (the original version of which dates back to the 1990s), while the base Encore and base Primo systems were next-generation catalog systems that debuted in 2007.

Activity: Vendor Investigations

After identification of existing and under development discovery services, a next step in UNLV's detailed vendor investigations included the creation of a uniform, comprehensive question list sent to each of the five vendors. The Discovery Task Force ultimately developed a list of 71 questions divided into nine functional areas, as follows, with an example question:

> Section 1: Background. "When did product development begin (month, year)?"
>
> Section 2: Locally Hosted Systems and Associated Metadata. "With what metadata schemas does your discovery platform work? (e.g., MARC, Dublin Core, EAD, etc.)"
>
> Section 3: Publisher/Aggregator Coverage (Full Text and Citation Content). "With approximately how many publishers/aggregators have you forged content agreements ?"
>
> Section 4: Records Maintenance and Rights Management. "How is your system initialized with the correct set of rights management information when a new library customer subscribes to your product?"

Section 5: Seamlessness & Interoperability with Existing Content Repositories. "For ILS records related to physical holdings, is status information provided directly within the discovery service results list?"

Section 6: Usability Philosophy. "Describe how your product incorporates published, established best practices in terms of a customer focused, usable interface."

Section 7: Local "Look & Feel" Customization Options. "Which of the following can the library control: Color Scheme; Logo / Branding; Facet Categories and placement; etc."

Section 8: User Experience (Presentation, Search Functionality, and What the User Can Do With the Results). "At what point does a user leave the context and confines of the discovery interface and enter the interface of a different system, whether remote or local?"

Section 9: Administration Module & Statistics. "Describe in detail the statistics reporting capabilities offered by your system. Does your system provide the following sets of statistics . . ."

All vendors were given 2–3 weeks to respond, and all vendors responded. It was evident from the uneven level of responses to the questions that the vendors were at different developmental states with their products. Some vendors were still 6–9 months away from initial public release; some were not even firm on when their service would enter release. It was also observed that some vendors were less explicit in the level of detail provided, reflective of, or in some cases perhaps regardless of, development state. A refined subset of the original 71 questions appears as a list of 40 questions in appendix F.

Apart from the detailed question list, various sets of free and licensed information on these discovery services are available online, and the task force sought to identify and digest the information. The *Charleston Advisor* has conducted interviews with several of the library web-scale discovery vendors on their products, including EBSCO,[15] Serials Solutions,[16] and Ex Libris.[17] These interviews, each around a dozen questions, ask the vendors to describe their product, how it differs from other products in the marketplace, and include questions on metadata and content—all important questions. An article by Ronda Rowe reviews Summon, EDS, and WorldCat Local, and provides some analysis of each product on the basis of content, user interface and searchability, pricing, and contract options.[18] It also provides a comparison of 24 product features provided by these three services, such as "search box can be embedded in any webpage," "local branding possible," and "supports social networking." A wide variety of archived webcasts, many provided by *Library Journal*, are available through free registration, and new webcasts are being offered at time of writing; these presentations to some degree touch on discussions with the discovery vendors, and are often moderated or include company representatives as part of the discussion group.[19] Several libraries have authored reports and presentations that, at least partially, discuss information on particular services gained through their evaluations, which include dialog with the vendors.[20] Vendors themselves each have a section on their corporate website devoted to their service. Information provided on these websites ranges from extremely brief to, in the case of WorldCat Local, very detailed and informative. In addition, much can be gained by "test-driving" live implementations. As such, a listing of vendor website addresses

providing more information as well as a list of sample, live implementations is provided in appendix G.

Activities: Vendor Visits and Content Overlap Analysis

Each of the five vendors visited the UNLV Libraries in spring 2010. Vendor visits all occurred within a nine-day span; visits were intentionally scheduled close to each other to keep things fresh in the minds of library staff, and such proximity would help with product comparisons. Vendor visits lasted approximately half a day, and each vendor visit often included the field or regional sales representative as well as a product manager or technical expert. Vendor visits included a demonstration and Q&A for all library staff as well as invited colleagues from other southern Nevada libraries, a meeting with the Discovery Task Force, and a meeting with technical staff at UNLV responsible for website design and application development and customization. Vendors were each given a uniform set of fourteen questions on topics to address during their visit; these appear in appendix H. Questions were divided into the broad topical areas of content coverage, end user interface and functionality, and staff "control" over the end user interface. On average, approximately 30–40 percent of the library staff attended the open vendor demo and Q & A session.

Shortly after the vendor visits, a content-overlap analysis comparing UNLV serials holdings with indexed content in the discovery service was sought from each vendor. Given that the amount of content indexed by each discovery service was growing (and continues to grow) extremely rapidly as new publisher and aggregator content agreements are signed, this content-overlap analysis was intentionally not sought at an earlier date. Some vendors were able to provide detailed coverage information against our existing journal titles (UNLV currently subscribes to approximately 20,000 e-journals and provides access to another 7,000+ open-access titles). For others, this was more difficult. Recognizing this, the head of Collection Development was asked to provide a list of the "top 100" journal titles for UNLV based on such factors as usage statistics and whether the title was a core title for part of the UNLV curriculum. The remaining vendors were able to provide content coverage information against this critical title list. Four of the five products had quite comprehensive coverage (more than 80 percent) of the UNLV Libraries' titles. While outside the scope of this article, "coverage" can mean different things for different services. Driven by the publisher agreements they are able to secure, some discovery services may have extensive coverage for particular titles (such as the full text, abstracts, author-supplied keywords, subject headings, etc.), whereas other services, while covering the same title, may have "thinner" metadata, such as basic citation information (article title, publication title, author, publication date, etc.). More discussion on this topic is present in the January 2011 *Library Technology Reports* on library web-scale discovery services.[21]

Activity: Product Development Tracking

One aspect of web-scale discovery services, and the next-generation discovery layers that preceded them, is a rapid enhancement cycle, especially when juxtaposed against the turnkey-style ILS system that dominated library automation for many years. As an example, minor enhancements are provided by Serials Solutions to Summon approximately every three to four weeks; provided by EBSCO to EBSCO Discovery Service approximately every three months; and

provided by Ex Libris to Primo/Primo Central approximately every three months. Many vendors unveil updates coinciding with annual library conferences, and 2010 was no exception. In late summer/early fall 2010, the Discovery Task Force had conference calls or onsite visits with several of the vendors with a focused discussion on new enhancements and changes to services as well as to obtain answers to any questions that arose since their last visit several months earlier. Since the vendor visits in spring 2010, each service had changed, and two services had unveiled significantly different and improved interfaces.

The Discovery Task Force's understanding of web-scale discovery services had expanded greatly since starting their work. Coordinated with the second series of vendor visits and discussions, an additional list of more than two dozen questions, recognizing this refined understanding, was sent to the majority of vendors. A portion of these questions are provided as part of the refined list of questions presented in appendix F. This second set of questions dealt with complex discussions of metadata quality, such as what level of content publishers and aggregators were providing for indexing purposes, e.g., full text, abstracts, table of contents, author-supplied keywords or subject headings, or particular citation and record fields), and also the vendor's stance on content neutrality, i.e., whether they are entering into exclusive agreements with publishers and aggregators, and, if the discovery service vendor is owned by a company involved with content, if that content is promoted or weighted more heavily in result sets. Other questions dealt with such topics as current install base counts and technical clarifications about how their service worked. In particular, the questions related to content were tricky for many (not all) of the vendors to address. Still, the Discovery Task Force was able to get a better understanding of how things worked in the evolving discovery environment. Combined with the internal library perspective and the early adopter references, information gathered from vendors provided the necessary data set to submit a recommendation with confidence.

Activity: Recommendation

By mid-fall 2010, the Discovery Task Force had conducted and had at their disposal a tremendous amount of research. Recognizing how quickly these services change and the fact that a cyclical evaluation could occur, the task force members felt they had met their charge. If all things failed during the next phase—implementation—at least no one would be able to question the thoroughness of the task force's efforts. Unlike the hasty decision, which in part led to a less than stellar experience with federated search a few years earlier, the evaluation process to recommend a new web-scale discovery service was deliberate, thorough, transparent, and vetted with library stakeholders.

Given the Discovery Task Force was entering its final phase, official price quotes were sought from each vendor. Each task force member was asked to develop a pro/con list for all five identified products based on the knowledge that was gained. These lists were anonymized and consolidated into a single, extensive pro/con list for each service. Some of the pros and cons were subjective (such as the interface aesthetics), some were objective (such as a particular discovery service not offering a desired feature). At one of the final meetings of the task force, members reaffirmed the three top contenders, indicated the other two were no longer under consideration and, afterward, were asked to rank their first, second, and third choices for the remaining services. While complete consensus wasn't achieved, there was a resounding first choice, second choice, and third

choice. The task force presented a summary of findings at a meeting open to all library staff. This meeting summarized the research and evaluation steps the task force had conducted over the past year, framed each of the three shortlisted services by discussing some strengths and weaknesses of each service as observed by the task force, and sought to answer any questions from the library at large. Prior to drafting the final report and making the recommendation to the dean of Libraries, several task force members led a discussion and final question and answer at a Libraries' cabinet meeting, one of the high-level administrative groups at the UNLV Libraries. Vetting by this body represented the last step related to the Discovery Task Force's investigation, evaluation, and recommendation for purchase of a library web-scale discovery service. The recommendation was broadly accepted by the Library cabinet, and shortly afterward the Discovery Task Force was officially disbanded, having met its goal of investigating, evaluating, and making a recommendation for purchase of a library web-scale discovery service.

Next Steps

The dialog above describes the research, evaluation, and recommendation model used by the UNLV Libraries to select a web-scale discovery service. Such a model and the associated appendixes could serve as a framework, with some adaptations perhaps, for other libraries considering the evaluation and purchase of a web-scale discovery service. Together, the Discovery Task Force's internal and external research and evaluation provided a substantive base of knowledge on which to make a recommendation. After its recommendation, the project progressed from a research and recommendation phase to an implementation phase. The Libraries' cabinet brainstormed a list of more than a dozen concise implementation bullet points—steps that would need to be addressed—including the harvesting and metadata mapping of local library resources, local branding and some level of customization work, and integration of the web-scale discovery search box in the appropriate locations on the Libraries' website. Project implementation co-managers were assigned (the director of Technical Services and the Web Technical Support manager), as well as key library personnel who would aid in one or more implementation steps. In January 2011, the implementation commenced, with an expected public launch of the new service planned for mid-2011. The success of a web-scale discovery service at the UNLV Libraries is a story yet to be written, but one full of promise.

Acknowledgements

REFERENCES

1. Marcia J. Bates, *Improving User Access to Library Catalog and Portal Information*, final report, version 3 (Washington, DC: Library of Congress, 2003), 4, http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf (accessed September 10, 2010).

2. Roger C. Schonfeld and Ross Housewright, *Faculty Survey 2009: Key Strategic Insights for Libraries, Publishers, and Societies* (New York: Ithaka S+R, 2010), 4, http://www.ithaka.org/ithaka-s-r/research/faculty-surveys-2000-2009/Faculty%20Study%202009.pdf (accessed September 10, 2010).

3. OCLC, *Online Catalogs: What Users and Librarians Want* (Dublin, OH: OCLC, 2009), 20, http://www.oclc.org/reports/onlinecatalogs/fullreport.pdf (accessed September 10, 2010).

4. Ibid, vi.

5. Ibid, 14.

6. Karen Calhoun, *The Changing Nature of the Catalog and Its Integration with Other Discovery Tools: Final Report* (Washington, DC: Library of Congress, 2006), 35, http://www.loc.gov/catdir/calhoun-report-final.pdf (accessed September 10, 2010).

7. Bibliographic Services Task Force, *Rethinking How We Provide Bibliographic Services for the University of California: Final Report* (**[Pub location?]** University of California Libraries, 2005), 2, http://libraries.universityofcalifornia.edu/sopag/BSTF/Final.pdf (accessed September 10, 2010).

8. OCLC, *College Students' Perceptions of Libraries and Information Resources* (Dublin, OH: OCLC, 2006), part 1, page 4, http://www.oclc.org/reports/pdfs/studentperceptions.pdf (accessed September 10, 2010).

9. Doug Way, "The Impact of Web-Scale Discovery on the Use of a Library Collection," *Serials Review*, in press.

10. Bill Kelm, "WorldCat Local Effects at Willamette University," presentation, Prezi, July 21, 2010, http://prezi.com/u84pzunpb0fa/worldcat-local-effects-at-wu/ (accessed Sept 10, 2010).

11. Michael Boock, Faye Chadwell, and Terry Reese, "WorldCat Local Task Force Report to LAMP,"March 27, 2009, http://hdl.handle.net/1957/11167 (accessed February 12, 2012).

12. Michael Boock et al., "Discovery Services Task Force Recommendation to University Librarian," http://hdl.handle.net/1957/13817 (accessed February 12, 2012).

13. Ken Varnum et al., "University of Michigan Library Article Discovery Working Group Final Report," Umich, January 29, 2010, http://www.lib.umich.edu/files/adwg/final-report.pdf.**[Access date?]**

14. Jennifer Ward, Pam Mofjeld, and Steve Shadle, "WorldCat Local at the University of Washington Libraries," *Library Technology Reports* 44, no. 6 (August/September 2008).

15. Dennis Brunning and George Machovec, "An Interview with Sam Brooks and Michael Gorrell on the EBSCOhost Integrated Search and EBSCO Discovery Service," *Charleston Advisor* 11, no. 3 (January 2010): 62–65.

16. Dennis Brunning and George Machovec, "Interview About Summon with Jane Burke, Vice President of Serials Solutions," *Charleston Advisor* 11, no. 4 (April 2010): 60–62.

17. Dennis Brunning and George Machovec, "An Interview with Nancy Dushkin, VP Discovery and Delivery Solutions at Ex Libris, Regarding Primo Central," *Charleston Advisor* 12, no. 2 (October 2010): 58–59.

18. Ronda Rowe, "Web-Scale Discovery: A Review of Summon, EBSCO Discovery Service, and WorldCat Local," *Charleston Advisor* 12, no. 1 (October 2010): 5–10.

19. *Library Journal* archived webcasts are available at http://www.libraryjournal.com/csp/cms/sites/LJ/Tools/Webcast/index.csp (accessed Sept 10, 2010).

20. Boock, Chadwell, and Reese, "WorldCat Local Task Force Report to LAMP"; Boock et al., "Discovery Services Task Force Recommendation to University Librarian"; Ken Varnum et al., "University of Michigan Library Article Discovery Working Group Final Report."

21. Jason Vaughan, "Library Web-scale Discovery Services," *Library Technology Reports* 47, no. 1 (January 2011).

Note: Appendices A–H available as supplemental files.

Appendices

Appendix A. Discovery Task Force Timeline

Appendix B. Discovery Task Force Charge

Appendix C. Discovery Task Force: Staff Survey 1 Questions

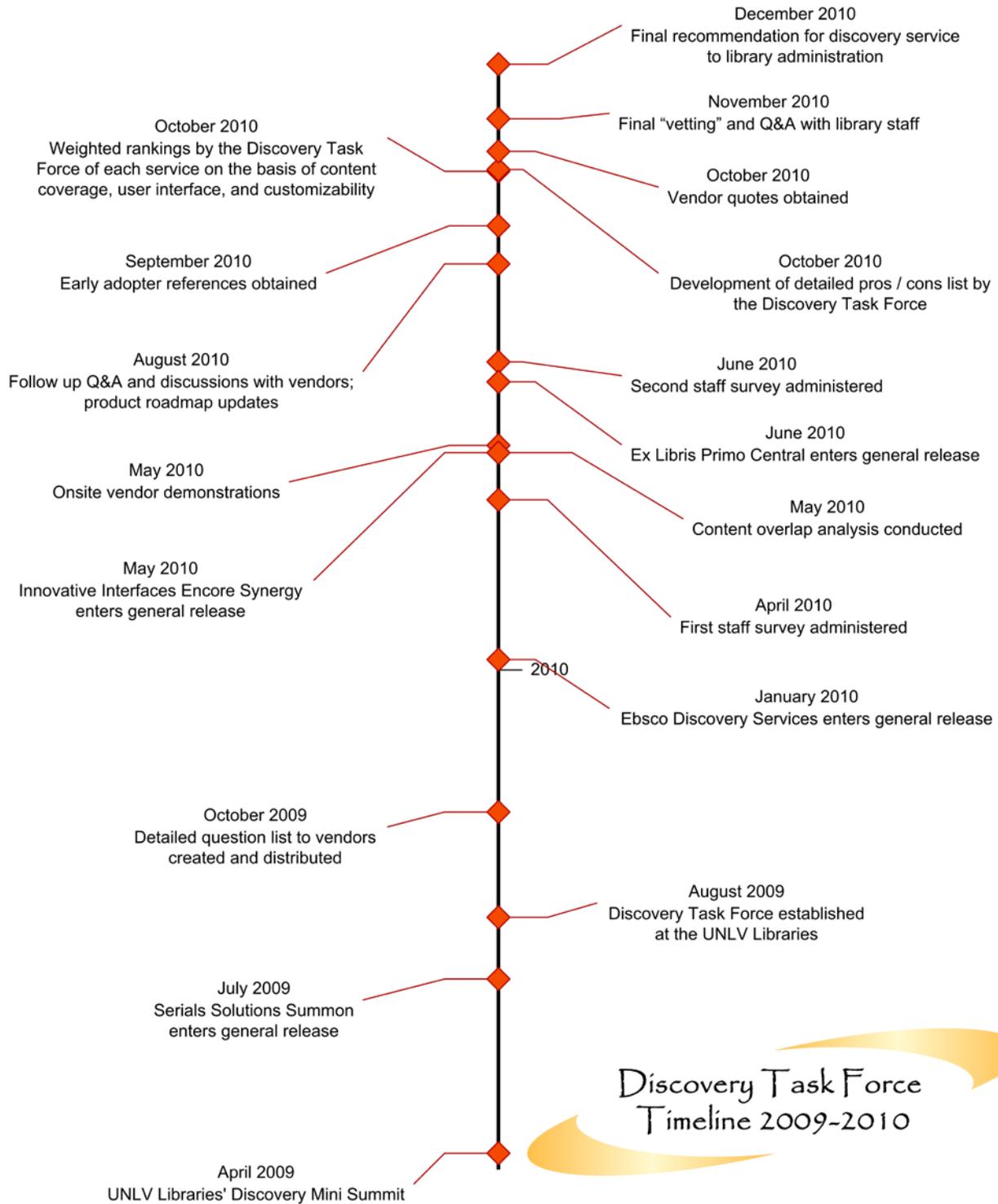Appendix D. Discovery Task Force: Staff Survey 2 Questions

Appendix E. Discovery Task Force: Early Adopter Questions

Appendix F. Discovery Task Force: Initial Vendor Investigation Questions

Appendix G. Vendor Websites and Example Implementations

Appendix H. Vendor Visit Questions

Appendix A. Discovery Task Force Timeline

December 2010
Final recommendation for discovery service
to library administration

November 2010
Final "vetting" and Q&A with library staff

October 2010
Weighted rankings by the Discovery Task
Force of each service on the basis of content
coverage, user interface, and customizability

October 2010
Vendor quotes obtained

September 2010
Early adopter references obtained

October 2010
Development of detailed pros / cons list by
the Discovery Task Force

August 2010
Follow up Q&A and discussions with vendors;
product roadmap updates

June 2010
Second staff survey administered

June 2010
Ex Libris Primo Central enters general release

May 2010
Onsite vendor demonstrations

May 2010
Content overlap analysis conducted

May 2010
Innovative Interfaces Encore Synergy
enters general release

April 2010
First staff survey administered

— 2010

January 2010
Ebsco Discovery Services enters general release

October 2009
Detailed question list to vendors
created and distributed

August 2009
Discovery Task Force established
at the UNLV Libraries

July 2009
Serials Solutions Summon
enters general release

*Discovery Task Force
Timeline 2009-2010*

April 2009
UNLV Libraries' Discovery Mini Summit

Appendix B. Discovery Task Force Charge

Discovery Task Force Charge

Informed through various efforts and research at the local and broader levels, and as expressed in the Libraries 2010/12 strategic plan, the UNLV Libraries have the desire to enable and maximize the discovery of library resources for our patrons. Specifically, the UNLV Libraries seek a unified solution which ideally could meet these guiding principles:

• Creates a unified search interface for users pulling together information from the library catalog as well as other resources (e.g. journal articles, images, archival materials).

• Enhances discoverability of as broad a spectrum of library resources as possible

• Intuitive: minimizes the skills, time, and effort needed by our users to discover resources

•Supports a high level of local customization (such as accommodation of branding and usability considerations)

• Supports a high level of interoperability (easily connecting and exchanging data with other systems that are part of our information infrastructure)

•Demonstrates commitment to sustainability and future enhancements

•Informed by preferred starting points

As such, the Discovery Task Force advises Libraries Administration on a solution that appears to best meet the goal of enabling and maximizing the discovery of library resources. A bulk of the work will entail a marketplace survey and evaluation of vendor offerings.

Charge

Specific deliverables for this work include:

1. Identify vendor next generation discovery platforms, whether established and currently on the market, or those publicized and at an advanced stage of development, with an expectation of availability within a year's time. Identify & create a representative list of other academic libraries which have implemented or purchased currently available products.

2. Create a checklist / criteria of functional requirements / desires for a next generation discovery platform.

3. Create lists of questions to distribute to potential vendors and existing customers of next generation discovery platforms. Questions will focus on broad categories such as the following:

   a. Seek to understand how content hosted in our current online systems (III catalog, CONTENTdm, locally created databases, vendor databases, etc.) could/would (or not be able

to) be incorporated or searchable within the discovery platform. Apart from our existing online systems as we know them today, the task force will explore, in general terms, how new information resources could be incorporated into the discovery platform. More explicitly, the task force will seek an understanding of what types of existing records are discoverable within the vendor's next generation discovery platform, and seek an understanding of what basic metadata must exist for an item to be discoverable.

   b. Seek to understand whether the solution relies on federated search, the creation of a central site index via metadata harvesting, or both, to enable discovery of items.

   c. Additional questions, such as pricing, maintenance, install base, etc.

4. Evaluate gathered information and seek feedback from library staff.

5. Provide to the Dean's Directs a final report which summarizes the task force findings. This report will include a recommended product(s) and a broad, as opposed to detailed, summary of workload implications related to implementation and ongoing maintenance. The final report should be provided to the Dean's Directs by February 15, 2010.

Boundaries

The work of the task force does not include:

• Detailing the contents of "hidden collections" within the Libraries and seeking to make a concrete determination that such hidden collections, in their current form, would be discoverable via the new system.

• Conducting an inventory, recommending, or prioritizing collections or items which should be cataloged or otherwise enriched with metadata to make them discoverable.

• Coordination with other southern Nevada NSHE entities.

• An ILS marketplace survey. The underlying Innovative Millennium System is not being reviewed for potential replacement.

• Implementation of a selected product.

[the charge concluded with a list of members for the Task Force]

Appendix C. Discovery Task Force: Staff Survey 1 Questions

"RANK" means the SurveyMonkey question will be set up such that each option can only be chosen once, and will be placed on a scale that corresponds to the number of choices overall.

"RATE" means there will be a 5 point Likert scale ranging from strongly disagree to strongly agree.

Section 1: Customization. The "Staff Side" of the House

1. Customization. It is important for the Library to be able to control/tweak/influence the following design element [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• General color scheme

• Ability to include a UNLV logo somewhere on the page.

• Ability to add other branding elements to the page.

• Ability to add one or more library specified links prominently in the interface (example: a link to the Libraries' home page)

• Able to customize the name of the product (meaning, the vendor's name for the product doesn't need to be used nor appear within the interface)

• Ability to embed the search box associated with the discovery platform elsewhere into the library website, such as the homepage (i.e. the user could start a search w/o having to directly go to the discovery platform

2. Customization. Are there any other design customization capabilities that are significantly important? Please list, and please indicate if this is a high, low, or medium priority in terms of importance to you. (freetext box)

3. Search Algorithms. It is important for the Library to be able to change or tweak the platform's native search algorithm to be able to promote desired items such that they appear higher in the returned list of [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

[e.g. The Library, at its option, could tweak one or more search algorithms to more heavily weight resources it wants to promote. For example, if a user searches for "Hoover Dam" the library could set a rule that would heavily weight and promote UNLV digital collection images for Hoover Dam – those results would appear on the first page of results].

4. Statistics. The following statistic is important to have for the discovery platform [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• Number of searches, by customizable timeframe

•Number of item or article level records accessed (that is, a user clicks on something in the returned list of results)

• Number of searches generating 0 results

• Number of items accessed by type

• Number of items accessed by provider of content (that is, number of articles from particular database/fulltext vendor

5. Statistics. What other statistics would you like to see a discovery platform provide and how important is this to you? (freetext box)

6. Staff Summary. Please RANK on a 1-3 scale how important the following elements are, with a "1" being most important, a "2" being 2nd most important, and a 3 being 3rd most important.

• Heavy customization capabilities as described in questions 1 & 2 above

• Ability to tweak search algorithms as described in question 3

• Ability for the system to natively provide detailed search stats such as described in question 4, 5.

Section 2. The "End User" Side of the House

7. Searching. Which of the following search options is preferable when a user begins their search [choose one]

• The system has a "Google-like" simple search box

• The system has a "Google-like" simple search box, but also has an advanced search capability (user can refine the search to certain categories: author, journal, etc.)

• No opinion

8. Zero Hit Searches. For a search that retrieves no actual results: [choose one]

• The system should suggest something else or ask, "Did you mean?"

• Retrieving precise results is more important and the system should not suggest something else or ask "Did you mean?"

• no opinion

9. De-duplication of similar items. Which of the following is preferable [choose one]

• The system automatically de-dupes records (the item only appears once in the returned list)

• The system does not de-dupe records (the same item could appear more than once in the returned list, such as when we have overlapping coverage of a particular journal from multiple subscription vendors)

• No opinion

10. Sorting of Returned Results. It is important for the user to be able to sort or reorder a list of returned results by . . [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• Publication Date

• Alphabetical by Author Name

• Alphabetical by Title

• Full Text Items First

• By Media Type (examples: journal, book, image, etc)

11. Web 2.0 Functionality on Returned Results. The following items are important for a discovery platform to have . . [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

(note, if necessary, please conduct a search in the Libraries' Encore system to help illustrate / remember some of the features/jargon mentioned below. In Encore, "Facets" appear on the left hand side of the screen; the results with book covers, "add to cart," and "export" features appear in the middle; and a tag cloud to the right. Note: this question is asking about having the particular feature regardless of which vendor, and not how well or how poorly you think the feature works for the Encore system)

• A tag cloud

• Faceted searching

• Ability to add user-generated tags to materials ("folksonomies")

• Ability for users to write and post a review of an item

• Other (please specify)

12. Enriched Record Information on Returned Results. The following items are important to have in the discovery system . . . [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• Book covers for items held by the Libraries

• A Google Books preview button for print items held by the Libraries

• Displays item status information for print items held by the Libraries (example: available, checked out)

13. What the User Can do With the Results. The following functionality is important to have in the discovery system . . [Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• Retrieve the fulltext of an item with only a single click on the item from the initial list of returned results

• Ability to add items to a cart for easy export (print, email, save, export to Refworks)

• Ability to place an InterLibrary Loan / LINK+ Request for an item

• System has a login/user account feature which can store user search information for later. In other words, a user could potentially log in to retrieve saved searches, previously stored items, or create alerts when new materials become available.

14. Miscellaneous. The following feature/attribute is important to have in the discovery system . . .

[Strongly Disagree / Disagree / Neither Agree or Disagree / Agree / Strongly Agree]

• The vendor has an existing mobile version of their discovery tool for use by smartphones or other small internet-enabled devices.

• The vendor has designed the product such that it can be incorporated into other sites used by students, such as WebCampus and/or social networking sites. Such "designs" may include the use of persistent URLs to embed hyperlinks, the ability to place the search box in another website, or specifically designed widgets developed by the vendor

• Indexing and availability of newly published items occurs within a matter of days as opposed to a week or perhaps a month.

• Library catalog authority record information is used to help return proper results and/or populate a tag cloud.

15. End User Summary. Please RANK on a 1-8 scale how important the following elements are; a "1" means you think it is the most important, a "2" second most important, etc.

• System offers a "Google-like" simple search box only, as detailed in question 7 above

• System offers a "did you mean?" or alternate suggestions for all searches retrieving 0 results as detailed in question 8 above (obviously, if you value precision of results over "did you mean" functionality, you would rank this toward the lower end of the spectrum).

• System de-dupes similar items as detailed in question 9 above(if you believe the system should not de-dupe similar items, you would rate this toward the lower end of the spectrum)

• System provides multiple sort options of returned results as detailed in question 10 above

• System offers a variety of Web 2.0 features as detailed in question 11 above

• System offer enriched record information as detailed in question 12 above

• System offers flexible options for what a user can do with the results, as detailed in question 13 above

• System has one or more miscellaneous features as detailed in question 14 above.

Section 3: Content

16. Incorporation of Different Information Types. In an ideal world, a discovery platform would incorporate ALL of our electronic resources, whether locally produced or licensed/purchased from vendors. Below is a listing of different information types. Please RANK on a scale of 1-10 how vital it is

that a discovery platform accommodate these information types ("1" is the most important item in your mind, a "2" is second most important, etc).

a. Innopac Millennium records for UNLV print & electronic holdings

b. LINK+ records for print holdings held within the LINK+ consortium

c. Innopac authority control records

d. Records within OCLC WorldCat

e. CONTENTdm records for digital collection materials

f. bePRESS Digital Commons Institutional Repository materials

g. Locally created Web accessible database records (e.g. the Special Collections & Architecture databases)

h. Electronic Reserves materials hosted in ERES

i. A majority of the citation records from non fulltext, vendor licensed online index/abstract/citation databases (e.g. The "Agricola" database)

 j. A majority of the fulltext articles or other research contained in many of our vendor licensed online resources (e.g. "Academic Search Premier" which contains a lot of full text content, and the other fulltext resource packages / journal titles we subscribe to)

17. LOCAL Content. Related to item (g) in the question immediately above, please list any locally produced collections that are currently available either on the website, or in electronic format as a word document, excel spreadsheet or access database (and not currently available on the website) that you would like the discovery platform to incorporate. (freetext box)

18. Particular Sets of Licensed Resources, What's Important? Please rank which of the licensed (full text or primarily full text) existing publishers below are most important for a discovery platform to accommodate.

Elsevier
Sage
Wiley
Springer
American Chemical Society
Taylor & Francis (Informaworld)
IEEE
American Institute of Physics
Oxford
Ovid
Nature
Emerald

Section 4: Survey Summary

19. Overarching Survey Question. The questions above were roughly categorized into three areas. Given that no discovery platform will be everything to everybody, please RANK on a 1-3 scale what the most important aspects of a discovery system are to you (1 is most critical, 2 is second in importance overall, etc.)

• The platform is highly customizable by staff (types of things in area 1 of the survey)

• The platform is highly flexible from the end-user standpoint (type of things in area 2 of the survey)

• The platform encompasses a large variety of our licensed and local resources (type of things in area 3 of the survey)

20. Additional Input. The survey above is roughly drawn from a larger list of 71 questions sent to the Discovery Task Force vendors. What other things do you think are REALLY important when thinking about a next-generation discovery platform? (freetext input, you may write a sentence or a book)

21. Demographic. What Library division do you belong to?

Library Administration

Library Technologies

Research & Education

Special Collections

Technical Services

User Services

Appendix D. Discovery Task Force: Staff Survey 2 Question

For the comparison questions, products are listed by order of vendor presentation. Please mark an answer for each product.

PART I. Licensed Publisher CONTENT (e.g. fulltext journal articles; citations / abstracts)

SA = Strongly Agree; A = Agree; N= Neither Agree nor Disagree; D = Disagree; SD = Strongly Disagree

1. "The Discovery Platform appears to ADEQUATELY cover a MAJORITY of the CRITICAL publisher titles."

	SA	A	N	D	SD	I don't know enough about the content coverage for this product to comment
Ex Libris Primo Central						
OCLC WorldCat Local						
Ebsco Discovery Services						
Innovative Encore Synergy						
Serials Solutions Summon						

2. "The Discovery Platform appears to ADEQUATELY cover a MAJORITY of the SECOND-TIER or SOMEWHAT LESS CRITICAL publisher titles."

	SA	A	N	D	SD	I don't know enough about the content coverage for this product to comment
Ex Libris Primo Central						
OCLC WorldCat Local						
Ebsco Discovery Services						
Innovative Encore Synergy						
Serials Solutions Summon						

3. Overall, from the CONTENT COVERAGE point of view, please rank each platform from best to worst.

	Worst	2nd Worst	Middle	2nd Best	Best
Ex Libris Primo Central					
OCLC WorldCat Local					
Ebsco Discovery Services					
Innovative Encore Synergy					
Serials Solutions Summon					

4. Regardless of a best to worst ranking, please indicate if the products were, overall, ACCEPTABLE or UNACCEPTABLE to you from the CONTENT COVERAGE standpoint.

	Unacceptable	Acceptable
Ex Libris Primo Central		

OCLC WorldCat Local		
Ebsco Discovery Services		
Innovative Encore Synergy		
Serials Solutions Summon		

PART II. END-USER FUNCTIONALITY & EASE OF USE

5. From the USER perspective, how functional do you think the discovery platform is? Are the facets and/or other methods that one can use to limit or refine a search appropriate? Were you satisfied with the export options offered by the system (email, export into Refworks, print, etc.)? If you think Web 2.0 technologies are important (tag cloud, etc.), were one or more of these present (and well executed) in this product?

	The platform appears to be SEVERELY limited in major aspects of end user functionality	The platform appears to have some level of useful functionality, but perhaps not as much or as well executed as some competing products.	Yes, the platform seems quite rich in terms of end user functionality, and such functions are well executed.	I can't comment on this particular product because I didn't see the vendor demo, haven't visited any of the live implementations linked on the discovery wiki page, or otherwise don't have enough information.
Ex Libris Primo Central				
OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore Synergy				
Serials Solutions Summon				

6. From the USER perspective, for a full-text pdf journal article, how EASY is it to retrieve the full-text? Does it take many clicks? Are there confusing choices?

	It's very cumbersome trying to retrieve the full text of an item, there are many clicks, and/or it's simply confusing when going through the steps to retrieve the full text.	It's somewhat straightforward to retrieve a full text item, but perhaps it's not as easy or as well executed as some of the competing products	It's quite easy to retrieve a full text item using this platform, as good as or better than the competition, and I don't feel it would be a barrier to a majority of our users.	I can't comment on this particular product because I didn't see the vendor demo, haven't visited any of the live implementations linked on the discovery wiki page, or otherwise don't have enough information.
Ex Libris Primo Central				

OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore Synergy				
Serials Solutions Summon				

7. How satisfied were you with the platform's handling of "dead end" or "zero hit" searches? Did the platform offer "did you mean" spelling suggestions? Did the platform offer you the option to request the item via doc delivery / LINK+?   Is the vendor's implementation of such features well executed, or were they difficult, confusing, or otherwise lacking?

	The platform appears to be severely limited in or otherwise poorly executes how it responds to a dead end or zero hit search.	The platform handled dead end or zero hit results, but perhaps not as seamlessly or as well executed as some of the competing products.	I was happy with how the platform handled "dead end" searches, and such functionality appears to be well executed, as good as or better than the competition.	I can't comment on this particular product because I didn't see the vendor demo, haven't visited any of the live implementations linked on the discovery wiki page, otherwise don't have enough information.
Ex Libris Primo Central				
OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore Synergy				
Serials Solutions Summon				

8. How satisfied were you with the platform's integration with the OPAC? Were important things such as call numbers, item status information, and enriched content immediately available and easily viewable from within the discovery platform interface, or did it require an extra click or two into the OPAC – and did you find this cumbersome or confusing?

	The platform provides minimal OPAC item information, and a user	The platform appeared to integrate ok with the OPAC in	I was happy with how the platform integrated with the	I can't comment on this particular product because I didn't see the

	would have to click through to the OPAC to get the information they might really need; and/or it took multiple clicks or was otherwise cumbersome to get the relevant item level information	terms of providing some level of relevant item level information, but perhaps not as much or as well executed as competing products.	OPAC. A majority of the OPAC information was available in the discovery platform, and/or their connection to the OPAC was quite elegant.	vendor demo, haven't visited any of the live implementations linked on the discovery wiki page, or otherwise don't have enough information.
Ex Libris Primo Central				
OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore Synergy				
Serials Solutions Summon				

9. Overall, from an END USER FUNCTIONALITY / EASE OF USE standpoint – how a user can refine a search, export results, easily retrieve the fulltext, easily see information from the OPAC record – please rank each platform from best to worst.

	Worst	2nd Worst	Middle	2nd Best	Best
Ex Libris Primo Central					
OCLC WorldCat Local					
Ebsco Discovery Services					
Innovative Encore Synergy					
Serials Solutions Summon					

10. Regardless of a best to worst ranking, please indicate if the products were, overall, ACCEPTABLE or UNACCEPTABLE to you from the USER FUNCTIONALITY / EASE OF USE standpoint.

	Unacceptable	Acceptable
Ex Libris Primo Central		
OCLC WorldCat Local		
Ebsco Discovery Services		
Innovative Encore Synergy		
Serials Solutions Summon		

PART III. STAFF CUSTOMIZATION

11. The "out of the box" design demo'ed at the presentation (or linked to the discovery wiki page – whichever particular implementation you liked best for that product) was . .

	Seriously lacking and I feel would need major design changes and customization by library Web technical staff.	Middle of the road – some things I liked, some things I didn't. The interface design was better than some competing products, worse than others.	Appeared very professional, clean, well organized, and usable; the appearance was better than most/all of the others products.	I can't comment on this particular product because I didn't see the vendor demo, haven't visited any of the live implementations linked on the discovery wiki page, or otherwise don't have enough information.
Ex Libris Primo Central				
OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore Synergy				
Serials Solutions Summon				

12. All products offer some level of customization options that allow at least SOME changes to the "out of the box" platform. Based on what the vendors indicated about the level of customization possible with the platform (e.g. look and feel, ability to add library links, ability to embed the search box on a homepage) do you feel there is enough flexibility with this platform for our needs?

	The platform appears to be severely limited in the degree or types of customization that can occur at the local level. We appear "stuck" with what the vendor gives us – for better or worse.	The platform appeared to have some level of customization, but perhaps not as much as some competing products.	Yes, the platform seems quite rich in terms of customization options under our local control; more so than the majority or all of the other products.	I can't comment on this particular product because I didn't see the vendor demo, don't have enough information, and/or would prefer to leave this question to technical staff to weigh in on.
Ex Libris Primo Central				
OCLC WorldCat Local				
Ebsco Discovery Services				
Innovative Encore				

Synergy				
Serials Solutions Summon				

13. Overall, from a STAFF CUSTOMIZATION standpoint – the ability to change the interface, embed links, define facet categories, define labels, place the searchbox in a different webpage, etc., please rank each platform from best to worst.

	Worst	2nd Worst	Middle	2nd Best	Best
Ex Libris Primo Central					
OCLC WorldCat Local					
Ebsco Discovery Services					
Innovative Encore Synergy					
Serials Solutions Summon					

14. Regardless of a best to worst ranking, please indicate if the products were, overall, ACCEPTABLE or UNACCEPTABLE to you from the STAFF CUSTOMIZATION standpoint.

	Unacceptable	Acceptable
Ex Libris Primo Central		
OCLC WorldCat Local		
Ebsco Discovery Services		
Innovative Encore Synergy		
Serials Solutions Summon		

PART IV. SUMMARY QUESTIONS

15. Overall, from a content coverage, user functionality, AND staff customization standpoint, please rank each product from best to worst.

	Worst	2nd Worst	Middle	2nd Best	Best
Ex Libris Primo Central					
OCLC WorldCat Local					
Ebsco Discovery Services					
Innovative Encore Synergy					
Serials Solutions Summon					

16. Regardless of a best to worst ranking, please indicate if the products were, overall, ACCEPTABLE or UNACCEPTABLE to you from the overall standpoint of content coverage, user functionality, AND staff customization standpoint.

	Unacceptable	Acceptable
Ex Libris Primo Central		
OCLC WorldCat Local		
Ebsco Discovery Services		
Innovative Encore Synergy		
Serials Solutions Summon		

PART V. ADDITIONAL THOUGHTS

17. Please share any additional thoughts you have on Ex Libris Primo Central. (freetext box)

18. Please share any additional thoughts you have on OCLC WorldCat Local. (freetext box)

19. Please share any additional thoughts you have on Ebsco Discovery Services. (freetext box)

20. Please share any additional thoughts you have on Innovative Encore Synergy. (freetext box)

21. Please share any additional thoughts you have on Serials Solutions Summon. (freetext box)

Appendix E. Discovery Task Force: Early Adopter Reference Questions

Author's note: Appendix E originally appeared in the January 2011 Library Technology Reports: Web Scale Discovery Services as chapter 7, "Questions to Consider."

Part 1 BACKGROUND

1. How long have you had your discovery service available to your end users? (what month and year did it become generally available to your primary user population, and linked to your public library website).

2. After you had selected a discovery service, approximately how long was the implementation period – how long did it take to "bring it up" for your end-users and make it available (even if in 'beta' form) on your library website?

3. What have you named your discovery service, and is it the 'default' search service on your website at this point? In other words, regardless of other discovery systems (ILS, Digital Collection Management System, IR, etc.), has the new discovery service become the default or primary search box on your website?

Part 2 CONTENT: Article Level Content Coverage & Scope

"Article Level Content" = articles from academic journals, articles from mainstream journals, newspaper content, conference proceedings, open access content

4. In terms of article level content, do you feel the preindexed, preharvested central index of the discovery platform adequately covers a majority of the titles important to your library's collection and focus?

5. Have you observed any particular strengths in terms of subject content in any of the three major overarching areas -- humanities, social sciences, sciences?

6. Have you observed any big, or appreciable, gaps in any of the three major overarching areas – humanities, social sciences, sciences?

7. Have you observed that the discovery service leans toward one or a few particular content **types** (e.g. peer reviewed academic journal content; mainstream journal content; newspaper article content; conference proceedings content; academic open access content)?

8. Are there particular publishers whose content is either not incorporated, (or not adequately incorporated), into the central index, that you'd like to see included (e.g. Elsevier journal content)?
9. Have you received any feedback, positive or negative, from your institution's faculty, related to the content coverage within the discovery service?

10. Taking all of the above questions into consideration, are you happy, satisfied, or dissatisfied with the scope of subject content, and formats covered, in the discovery platform's central index?

11. In general, are you happy with the level of article level metadata associated with the returned

citation level results (that is, before one retrieves the complete full text). In other words, the product may incorporate basic citation level metadata (e.g. title, author, publication info), or it may include additional enrichment content, such as abstracts, author supplied keywords, etc. Overall, how happy do you sense your library staff is with the quality and amount of metadata provided for a "majority" of the article level content indexed in the system?

Part 3 CONTENT: Your Local Library Resources

12. It's presumed that your local library ILS bib records have been harvested into the discovery solution. Do you have any other local "homegrown" collections – hosted by other systems at your library or institution – whose content has been harvested into the discovery solution? Examples would include digital collection content, institutional repository content, library subject guide content, or other specialized, homegrown local database content. If so, please briefly describe the content – focus of collection, type of content (images, articles, etc.), and a ballpark number of items. If no local collections other than ILS bib record content have been harvested, please skip to question 15.

13. [For local collections other than ILS Bib Records]. Did you use existing, vendor provided ingestors to harvest the local record content (i.e. ingestors to transfer the record content, apply any transformations and normalizations to migrate the local content to the underlying discovery platform schema)? Or did you develop your own ingestors from scratch, or using a toolkit or application profile template provided by the vendor?

14. [For local collections other than ILS Bib Records]. Did you need extensive assistance from the discovery platform vendor to help harvest any of your local collections into the discovery index? If so, regardless of whether the vendor offered this assistance for free or charged a fee, were you happy with the level of service received from the vendor?

15. Do you feel your local content (including ILS Bib records) is adequately "exposed" during a majority of searches? In other words, if your local harvested content equaled a million records, and the overall size of the discovery platform index was a hundred million records, do you feel your local content is "lost" for a majority of end user searches, or adequately exposed?

Part 4 INTERFACE: General Satisfaction Level

16. Overall, how satisfied are you and your local library colleagues with the discovery service's interface?

17. Do you have any sense of how satisfied faculty at your institution are with the discovery service's interface? Have you received any positive or negative comments from faculty related to the interface?

18. Do you have any sense of how satisfied your (non-faculty) end-users are with the discovery service's interface? Have you received any positive or negative comments from users related to the interface?

19. Have you conducted any end-user usability testing related to the discovery service? If so, can you provide the results, or otherwise some general comments on the results of these tests?

20. Related to searching, are you happy with the relevance of results returned by the discovery service? Have you noticed any consistent "goofiness," or surprises with the returned results? If you could make a

change in the relevancy arena, what would it be, if anything?

Part 5 INTERFACE: Local Customization

21. Has your library performed what you might consider any "major customization" to the product? Or has it primarily been customizations such as naming the service, defining hyperlinks and the color scheme? If you've done more extensive customization, could you please briefly describe, and was the product architecture flexible enough to allow you to do what you wanted to do (also see question 22 below, which is related).

22. Is there any particular feature or function that is missing or non-configurable within the discovery service that you wish were available?

23. In general, are you happy with the "openness" or "flexibility" of the system in terms of how customizable it is by your library staff?

Part 6: FINAL THOUGHTS

24. Overall, do you feel your selection of this vendor's product was a good one? Do you sense that your users – students and faculty – have positively received the product?

25. Have you conducted any statistics review or analysis (through the discovery service statistics, or link resolver statistics, etc.) that would indicate or at least suggest that the discovery service has improved the discoverability of some of your materials (whether local library materials or remotely hosted publisher content).

26. If you have some sense of the competition in the vendor discovery marketplace, do you feel this product offers something above and beyond the other competitors in the marketplace? If so, what attracted you to this particular product, what made it stand out?

Appendix F. Discovery Task Force: Initial Vendor Investigation Questions

Section 1: General / Background Questions

1. Customer Install Base

How many current customers do you have that have which have implemented the product at their institution? (the tool is currently available to users / researchers at that institution)

How many additional customers have committed to the product?

How many of these customers fall within our library type (e.g. higher ed academic, public, K-12)?

2. References

Can you provide website addresses for live implementations which you feel serve as a representative model matching our library type?

Can you provide references – the name and contact information for the lead individuals you worked with at several representative customer sites which match our library type?

3. Pricing Model, Optional Products

Describe your pricing model for a library type such as ours, including initial upfront costs and ongoing costs related to the subscription and technical support.

What optional add-on services or modules (federated search, recommender services, enrichment services) do you market which we should be aware of, related to and able to be integrated with your web scale discovery solution?

4. Technical Support and Troubleshooting

Briefly describe options customers have, and hours of availability, for reporting mission critical problems; and for reporting observed non mission-critical glitches.

Briefly describe any consulting services you may provide above and beyond support services offered as part of the ongoing subscription. (e.g. consulting services related to harvesting of a unique library resource for which an ingest/transform/normalize routine does not already exist).

Is there a process for suggesting enhancement requests for potential future incorporation into the product?

5. Size of the Centralized Index. How many periodical titles does your preharvested, centralized index encompass? How many indexed items?

6. Statistics.

Please describe what you feel are some of the more significant use, management or content related statistics available out-of-the-box with your system.

Are the statistics COUNTER compliant?

7. Ongoing Maintenance Activities, Local Library Staff. For instances where the interface and discovery service is hosted on your end, please describe any ongoing local library maintenance activities associated with maintaining the service for the local library's clientele (e.g. maintenance of the link resolver database; ongoing maintenance associated with periodic local resource harvest updates; etc.)

Section 2: Local Library Resources

8. Metadata Requirements and Existing Ingestors.

What mandatory record fields for a local resource has to exist for the content to be indexed and discoverable within your platform (title, date)?

Please verify that your platform has existing connectors -- ingest/transform/normalize tools and transfer mechanisms and/or application profiles for the following schema used by local systems at our library (e.g. MARC 21 bibliographic records; Unqualified / Qualified Dublin Core, EAD, etc.)

Please describe any standard tools your discovery platform may offer to assist local staff in crosswalking between the local library database schema and the underlying schema within your platform.

Our Library uses the ABC digital collection management software. Do you have any existing customers who also utilize this platform, whose digital collections have been harvested and are now exposed in their instance of the discovery product?

Our Library uses the ABC institutional repository software. Do you have any existing customers who also utilize this platform, whose digital collections have been harvested and are now exposed in their instance of the discovery product?

9. Resource Normalization. Is content for both local and remote content normalized to a single schema? If so, please offer comments on how local and remote (publisher/aggregator) content is normalized to this single underling schema. To what degree can collections from different sources have their own unique field information which is displayed and/or figures into the relevancy ranking algorithm for retrieval purposes.

10. Schedule.

For records hosted in systems at the local library, how often do you harvest information to account for record updates, modifications, deletions?

Can the local library invoke a manual harvest of locally hosted resource records on a per-resource basis (e.g. from a selected resource – for example, if the library launches a new digital collection and want the records to be available in the new discovery platform shortly after they are available in our local digital collection management system, is there a mechanism to force a harvest prior to the next regularly scheduled harvest routine?

After harvesting, how long does it typically take for such updates, additions, and deletions to be reflected in the searchable central index?

11. Policies / Procedures. Please describe any general policies and procedures not already addressed which the local library should be aware of as relates to the harvesting of local resources.

12. Consortial Union Catalogs. Can your service harvest or provide access to items within a consortial or otherwise shared catalog (e.g. the INN-REACH catalog). Please describe.

Section 3: Publisher and Aggregator Indexed Content

13. Publisher/Aggregator Agreements: General

With approximately how many publishers have you forged content agreements with?

Are these agreements indefinite or do they have expiration dates?

Have you entered into any exclusive agreements with any publishers/aggregators (i.e. the publisher/aggregator is disallowed from forging agreements with competing discovery platform vendors, or disallowed from providing the same deep level of metadata/full text for indexing purposes).

14. Comments on Metadata Provided. Could you please provide some general comments on the level of data provided to you, for indexing purposes, by the "majority" of major publishers/aggregators with which you have forged agreements. Please describe to what degree the following elements play a role in your discovery service:

    a. "Basic" bibliographic information (article title/journal title/author/publication information)

    b. Subject descriptors

    c. Keywords (author supplied?)

    d. Abstracts (author supplied?)

    e. Full text

15. Topical Content Strength

Do you feel there is a particular content area that you feel the service covers especially well or leans heavily toward (e.g. Humanities, Social Sciences, Sciences).

Do you feel there is a particular content type that you feel the service covers very well or leans heavily toward (scholarly journal content, mainstream journal content, newspapers, conference proceedings).

What subject / content areas, if any, do you feel the service may be somewhat weak? Are there current efforts to mitigate these weaknesses (e.g. future publisher agreements on the horizon)?

16. Major Publisher Content Agreements. Are there major publisher agreements that you feel are especially significant for your service? If so, which publishers, and why (e.g. other discovery platform vendors may not have such agreements with those particular providers; the amount of content was so great that it greatly augmented the size and scope of your service; etc.)

17. Content Considered Key by Local Library (by publisher). Following is a list of some major publishers whose content the library licenses which is considered "key." Has your company forged agreements with these publishers to harvest their materials. If so please describe in general the scope of the agreement. How many titles are covered for each publisher? What level of metadata are they providing to you for indexing purposes (e.g. basic citation level metadata – title, author, publication date; abstracts; full text).

    A. ex. Elsevier

    B. ex. Sage

    C. ex. Taylor and Francis

    D. ex. Wiley / Blackwell

18. Content Considered Key by Local Library (by title). Following is a list of some major journal / newspaper titles whose content the library licenses which is considered "key." Could you please indicate if your central index includes these titles, and if so, the level of indexing (e.g. basic citation level metadata – title, author, publication date; abstracts; full text).

    A. ex. Nature

    B. ex. American Historical Review

    C. ex. JAMA

    D. ex. Wall Street Journal

19. Google Books / Google Scholar. Do any agreements exist at this time to harvest the data associated with the Google Books or Google Scholar projects into your central index? If so, could you please describe the level of indexing (e.g. basic citation level metadata – title, author, publication date; abstracts; full text).

20. Worldcat Catalog. Does your service include the OCLC WorldCat catalog records? If so, what level of information is included? The complete record? Holdings information?

21. E-Book Vendors. Does your service include items from major e-book vendors?

22. Record Information. Given the fact that the same content (e.g. metadata for a unique article) can be provided by multiple sources (e.g. the original publisher of the journal itself, an open access repository, a database / aggregator, another database / aggregator, etc.), please provide some general comments on how records are built within your discovery service. For example:

A. You have an agreement with a particular publisher/aggregator and they agree to provide you with rich metadata for their content, perhaps even provide you with indexing they've already done for their content, and may even provide you with the full text for you to be able to "deep index" their content.

B. You've got an agreement with a particular publisher who happens to be the ONLY publisher/provider of that content. They may provide you rich info, or they may provide you rather weak info. In any case, you choose to incorporate this into your service, as they are the only provider/publisher of the info. Or,

alternately, they may not be the only publisher/provider of the info, but they are the only publisher/provider you've currently entered into an agreement with for that content.

C. For some items appearing within your service, content for those items is provided by multiple different sources whom you've made agreements with. In short, there will be in some/many cases of overlap for unique items, such as a particular article title. In such cases, do you create a "merged/composite/super record" -- where your service utilizes particular metadata from each of the multiple sources, creating a "strong" single record built from these multiple resources.

23. Deduping. Related to the question immediately above, please describe your services' approach (or not) to deduplicating items in your central index. If your service incorporates content for a same unique item from more than one content provider, does your index retrieve and display multiple instances of the same title? Or do you create a merged/composite/super record, and only this single record is displayed? Please describe.

Section 4: Open Access Content

24. Open Access Content Sources. Does your service automatically include (out of the box, no additional charge) materials from open access repositories? If so, could you please list some of the major repositories included (e.g. arXiv E-prints; Hindawi Publishing Corporation; the Directory of Open Access Journals; Hathi Trust Materials; etc.).

25. Open Access Content Sources: Future Plans. In addition to the current open access repositories that may be included in your service, are there other repositories whose content you are planning to incorporate in the future?

26. Exposure to other Libraries' Bibliographic / Digital Collection / IR Content. Are ILS bibliographic records from other customers using your discovery platform exposed for discoverability in the searchable discovery instance of another customer? Are digital collection records? Institutional repository records?

Section 5: Relevancy Ranking

27. Relevancy Determination. Please describe some of the factors which comprise the determination of relevancy within your service. What elements play a role, and how heavily are they weighted for purposes of determining relevancy?

28. Currency. Please comment on how heavily currency of an item plays in relevancy determination. Does currency weigh more heavily for certain content types (e.g. newspapers)?

29. Local Library Influence. Does the local library have any influence or level of control over the relevancy algorithm? Can they choose to "bump up" particular items for a search? Please describe.

30. Local Collection Visibility. Could you please offer some comments on how local content (e.g. ILS bibliographic records; digital collections) remains visible and discoverable within the larger pool of content indexed by your service? For example, local content may measures a million items, and your centralized index may cover half a billion items.

31. Exposure of Items with Minimal Metadata. Some items likely have lesser metadata than other items. Could you please offer some comments on how your system ensures discoverability for items with lesser or minimal metadata.

32. Full Text Searching. Does your service offer the capability for the user to search the fulltext of materials in your service (i.e. are they searching a full text keyword index?) If so, approximately what percentage of items within your service are "deep indexed?"

33. Please describe how your system deals when no hits are retrieved for a search. Does your system enable "best-match" retrieval – that is, something will always be returned or recommended? What elements play into this determination; how is the user prevented from having a completely "dead-end" search?

Section 6: Authentication and Rights Management

34. Open / Closed Nature of Your Discovery Solution. Does your system offer an unauthenticated view / access? Please describe and offer some comments on what materials will not be discoverable/visible for an unauthenticated user.

    A. Licensed Full Text

    B. Records specifically or solely sourced from Abstract and Indexing Databases

    C. Full citation information (e.g. an unauthenticated user may see just a title; an authenticated

    user would see fuller citation information)

    D. Enrichment information (such as book image covers, table of contents, abstracts, etc.)

    E. Other

35. Exposure of non-licensed Resource Metadata.

If one **weren't** to consider and take into account ANY e-journal/publisher package/database subscriptions & licenses the local library pays for, is there a base index of citation information that's exposed and available to all subscribers of your discovery service? This may include open access materials, and/or bibliographic information for some publisher / aggregator content (which often requires a local library license to access the full text). Please describe.

Would a user need to be authenticated to search (and retrieve results from) this "base index?"

Approximately how large is this "base index" which all customers may search, regardless of local library publisher/aggregator subscriptions.

36. Rights Management.

Please discuss how rights management is initialized and maintained in your system, for purposes of determining whether a local library user should have access to the full text (or otherwise "full resolution" if a library doesn't license the fulltext – such as resolution to a detailed citation/abstract).

Our library uses the ABC link resolver. Our library uses the ABC A-Z journal listing service. Our library uses the ABC electronic resource management system. Is your discovery solution compatible with one/all of these systems for rights management purposes? Is one approach preferable to the other, or does your approach explicitly depend on one of these particular services?

Section 7: User Interface

37. Openness to Local Library Customization. Please describe how "open" your system is to local library customization. For example, please comment on the local library's ability to

A. Rename the service

B. Customize the header and footer hyperlinks / color scheme

C. Choose which facet clusters appear

D. Define new facet clusters

E. Embed the search box in other venues

F. Create canned, pre-customized searches for an instance of the search box

G. Define and promote a collection, database, or item such that it appears at the top or on the first page of any search

I. Develop custom "widgits" offering extra functionality or download "widgits" from an existing user community (e.g. image retrieval widgits such as Flickr integration; library subject guide widgits such as Libguides integration; etc.

J. Incorporate links to external enriched content (e.g. Google Book Previews; Amazon.com item information)

K. Other

38. Web 2.0 Social Community Features. Please describe some current web 2.0 social features present in your discovery interface (e.g. user tagging, ratings, reviews, etc.). What, if any, plans do you have to offer or expand such functionality in future releases?

39. User Accounts.

Does your system offer user accounts?

If so, are these mandatory or optional?

What services does this user account provide?

A. Save a list of results to return to at a later time?

B. Save canned queries for later searching?

C. See a list of recently viewed items?

D. Perform typical ILS functions such as viewing checked out items / renewals / holds?

E. Create customized RSS feeds for a search

40. Mobile Interface. Please describe the mobile interfaces available for your product. Is it a browser based interface optimized for smallscreen devices? Is it a dedicated iPhone, Android, or Blackberry based executable application?

41. Usability Testing.

Briefly describe how your product incorporates published, established "best practices" in terms of a customer focused, usable interface. What usability testing have your performed and/or do you conduct on an ongoing basis?

Have any other customers that have gone live with your service completed usability testing that you're aware of?

Appendix G: Vendor Websites and Example Implementations

OCLC WorldCat Local

www.oclc.org/us/en/worldcatlocal/default.htm

Example Implementations:

Lincoln Trails Library System
www.lincolntrail.info/linc.html

University of Delaware
www.lib.udel.edu

University of Washington
www.lib.washington.edu

Willamette University
http://library.willamette.edu

Serials Solutions Summon

www.serialssolutions.com/summon

Example Implementations:

Dartmouth College
www.dartmouth.edu/~library/home/find/summon

Drexel University
www.library.drexel.edu

University of Calgary
http://library.ucalgary.ca

Western Michigan University
http://wmich.summon.serialssolutions.com

Ebsco Discovery Services

www.ebscohost.com/discovery

Example Implementations:

James Madison University
www.lib.jmu.edu

Mississippi State University
http://library.msstate.edu

Northeastern University
www.lib.neu.edu

University of Oklahoma
http://libraries.ou.edu

Innovative Interfaces Encore Synergy

encoreforlibraries.com/tag/encore-synergy

Example Implementations:

University of Nebraska-Lincoln
http://encore.unl.edu/iii/encore/home?lang=eng

University of San Diego
http://sallypro.sandiego.edu/iii/encore/home?lang=eng

Scottsdale Public Library
http://encore.scottsdaleaz.gov/iii/encore/home?lang=eng

Sacramento Public Library
http://find.saclibrarycatalog.org/iii/encore/home?lang=eng

Ex Libris Primo Central

www.exlibrisgroup.com/category/PrimoCentral

Example Implementations:

(Note: Example implementations are listed in alphabetical order. Some implementations are more open to search by an external audience, based on configuration decisions at the local library level.)

Brigham Young University ScholarSearch
www.lib.byu.edu

(Note: Choose All-in-One Search)

Northwestern University
http://search.library.northwestern.edu

Vanderbilt University DiscoverLibrary
http://discoverlibrary.vanderbilt.edu

(Note: Choose Books, Media, and More)

Yonsei University (Korea) WiSearch: Articles + Library Holdings
http://library.yonsei.ac.kr/main/main.do

(Note: Choose the Articles + Library Holdings link. The interface is available in both Korean and English; to change to English, select English at the top right of the screen after you have conducted a search and are within the Primo Central interface)

Appendix H. Vendor Visit Questions

Content

1. Please speak to how well you feel your product stacks up against the competition in terms of the LICENSED full-text / citation content covered by your product. Based on whatever marketplace or other competitive analysis you may have done, do you feel the agreements you've made with publishers equal, exceed, or trail the agreements other competitors have made?

2. From the perspective of an academic library serving undergraduate and graduate students as well as faculty, do you feel that there are particular licensed content areas your product covers very well (e.g. humanities, social sciences, sciences). Do you feel there are areas which you need to build up?

3. What's your philosophy going forward in inking future agreements with publishers to cover more licensed content? Are there particular key publishers your index currently doesn't include, but whom you are in active negotiations with?

4. We have several local content repositories, such as our digital collections in CONTENTdm, our growing IR repository housed in bePress, and locally developed, web-searchable mySQL databases. Given the fact that most discovery platforms are quite new, do you already have existing customers harvesting their local collections, such as the above, into the discovery platform? Have any particular, common problems surfaced in their attempts to get their local collections searchable and exposed in the discovery platform?

5. Let's say the library subscribes to an ejournal title – <u>Journal of Animal Studies</u> -- that's from a publisher with whom you don't have an agreement for their metadata, and thus, supposedly, don't index. If a student tried to search for an article in this journal – "Giraffe Behavior During the Drought Season," what would happen? Is this content still somehow indexed in your tool? Would the discovery platform invoke our link resolver? Please describe.

6. Our focus is your next generation discovery platform, and NOT on your "traditional" federated search product which may be able to cover other resources not yet indexed in your next generation discovery platform. That said, please BRIEFLY describe the role of your federated search product vis a vis the next generation discovery platform. Do you see your federated search product "going away" once more and more content is eventually indexed in your next generation discovery platform?

End User Interface & Functionality

7. Are there any particular or unique LOOK and FEEL aspects of your interface that you feel elevate your product above your competitors? If so, please describe.

8. Are there any particular or unique FUNCTIONALITY aspects of your product that you feel elevate it above the competition (e.g. presearch or postsearch refinement categories, export options, etc.)

9. Studies show that end users want very quick access to full text materials such as electronic journal articles and ebooks. What is your product's philosophy in regards to this? Does your platform, in your opinion, provide seamless, quick access to full text materials, with a minimum of confusion? Please describe.

Related to this, does your platform de-dupe results, or is the user presented with a list of choices for a single, particular journal article they are trying to retrieve?

In addition, please describe a bit how your relevancy ranking works for returned results. What makes an item appear first or on the first page of results?

10. Please describe how "well" your product integrates with the library's OPAC (in our case, Innovative's Millennium OPAC). What information about OPAC holdings can be viewed directly in the discovery platform w/o clicking into the catalog and opening a new screen (e.g. call #, availability, enriched content such as table of contents or book covers?)

In addition, our OPAC uses "scopes" which allow a user – if they choose – to limit at an outset (prior to a search being conducted) what

collection they are searching. In other words, these scopes are location based, not media type based. For our institution, we have a scope for the main library, one for each of our three branch libraries, and a scope for the entire UNLV collection. Would your system be able to incorporate or integrate these pre-existing scopes in an advanced search mode? And/or, could these location based scopes appear as facets which a user could use to drill down a results list?

11. What is your platform's philosophy in terms of "dead end searches." Does such a thing exist with your product? Please describe what happens if a user

a.) misspells a word

b.) searches for a book or journal title / article that our library doesn't own/license, but that we could acquire through interlibrary loan.

Staff "Control" over the End User Interface

12. How "open" is your platform to customization or interface design tweaks desired by the library? Are there any particular aspects that the library can customize with your product that you feel elevate it above your competitors (e.g. defining facet categories; completely redesigning the end-user interface with colors, links, logos; etc.)? What are the major things customizable by the library, and why do you think this is something important that your product offers.

13. How "open" is your platform to porting over to other access points? In other words, provided appropriate technical skills exist, can we easily embed the search box for your product into a different webpage? Could we create a "smaller," more streamlined version of your interface for smartphone access?

Overarching Question

14. In summary, what are some of the chief differentiators of your product from the competition? Why is your product the best and most worthy of serious consideration?