

Evaluating Web-Scale Discovery: A Step-by-Step Guide

Joseph Deodato

ABSTRACT

Selecting a web-scale discovery service is a large and important undertaking that involves a significant investment of time, staff, and resources. Finding the right match begins with a thorough and carefully planned evaluation process. To be successful, this process should be inclusive, goal-oriented, data-driven, user-centered, and transparent. The following article offers a step-by-step guide for developing a web-scale discovery evaluation plan rooted in these five key principles based on best practices synthesized from the literature as well as the author's own experiences coordinating the evaluation process at Rutgers University. The goal is to offer academic libraries that are considering acquiring a web-scale discovery service a blueprint for planning a structured and comprehensive evaluation process.

INTRODUCTION

As the volume and variety of information resources continue to multiply, the library search environment has become increasingly fragmented. Instead of providing a unified, central point of access to its collections, the library offers an assortment of pathways to disparate silos of information. To the seasoned researcher familiar with these resources and experienced with a variety of search tools and strategies, this maze of options may be easy to navigate. But for the novice user who is less accustomed to these tools and even less attuned to the idiosyncrasies of each one's own unique interface, the sheer amount of choice can be overwhelming. Even if the user manages to find their way to the appropriate resource, figuring out how to use it effectively becomes yet another challenge. This is at least partly due to the fact that the expectations and behaviors of today's library users have been profoundly shaped by their experiences on the web. Popular sites like Google and Amazon offer simple, intuitive interfaces that search across a wide range of content to deliver immediate, relevant, and useful results. In comparison, library search interfaces often appear antiquated, confusing, and cumbersome. As a result, users are increasingly relying on information sources that they know to be of inferior quality, but are simply easier to find. As Luther and Kelly note, the biggest challenge academic libraries face in today's abundant but fragmented information landscape is "to offer an experience that has the simplicity of Google—which users expect—while searching the library's rich digital and print collections—which users need."¹ In an effort to better serve the needs of these users and improve access to library content, libraries have begun turning to new technologies capable of providing deep discovery of their vast scholarly collections from a single, easy-to-use interface. These technologies are known as web-scale discovery services.

Joseph Deodato (jdeodato@rutgers.edu) is Digital User Services Librarian at Rutgers University, New Brunswick, New Jersey.



To paraphrase Hoepfner, a web-scale discovery service is a large central index paired with a richly featured user interface providing a single point of access to the library's local, open access, and subscription collections.² Unlike federated search, which broadcasts queries in real-time to multiple indexes and merges the retrieved results into a single set, web-scale discovery relies on a central index of preharvested data. Discovery vendors contract with content providers to index their metadata and full-text content, which is combined with the library's own local collections and made accessible via a unified index. This approach allows for rapid search, retrieval, and ranking of a broad range of content within a single interface, including materials from the library's catalog, licensed databases, institutional repository, and digital collections. Web-scale discovery services also offer a variety of features and functionality that users have come to expect from modern search tools. Features such as autocorrect, relevance ranking, and faceted browsing make it easier for users to locate library materials more efficiently while enhanced content such as cover images, ratings, and reviews offer an enriched user experience while providing useful contextual information for evaluating results.

Commercial discovery products entered the market in 2007 at a time when academic libraries were feeling pressure to compete with newer and more efficient search tools like Google Scholar. To improve the library search experience and stem the seemingly rising tide of defecting users, academic libraries were quick to adopt discovery solutions that promised improved access and increased usage of their collections. Yet despite the significant impact these technologies have on staff and users, libraries have not always undertaken a formal evaluation process when selecting a discovery product. Some were early adopters that selected a product at a time when there few other options existed on the market. Others served as beta sites for particular vendors or simply chose the product offered by their existing ILS or federated search provider. Still others had a selection decision made for them by their library director or consortium. However, despite rapid adoption, the web-scale discovery market has only just begun to mature. As products emerge from their initial release and more information about them becomes available, the library community has gained a better understanding of how web-scale discovery services work and their particular strengths and weaknesses. In fact, some libraries that have already implemented a discovery service are currently considering switching products. Whether your library is new to the discovery marketplace or poised for reentry, this article is intended to help you navigate to the best product to meet the needs of your institution. It covers the entire process from soup to nuts from conducting product research and drafting organizational requirements to setting up local trials and coordinating user testing. By combining guiding principles with practical examples, this article aims to offer an evaluation model rooted in best practices that can be adapted by other academic libraries.

LITERATURE REVIEW

As the adoption of web-scale discovery services continues to rise, a growing body of literature has emerged to help librarians evaluate and select the right product. Moore and Greene provide a

useful review of this literature summarizing key trends such as the timeframe for evaluation, the type of staff involved, the products being evaluated, and the methods and criteria used by evaluators.³ Much of the early literature on this subject focuses on comparisons of product features and functionality. Rowe, for example, offers comparative reviews of leading commercial services on the basis of criteria such as content, user interface, pricing, and contract options.⁴ Yang and Wagner compare commercial and open source discovery tools using a checklist of user interface features that includes search options, faceted navigation, result ranking, and Web 2.0 features.⁵ Vaughan provides an in-depth look at discovery services that includes an introduction to key concepts, detailed profiles on each major service provider, and a list of questions to consider when selecting a product.⁶ A number of authors have provided useful lists of criteria to help guide product evaluations. Hoepfner, for example, offers a list of key factors such as breadth and depth of indexing, search and refinement options, branding and customization, and tools for saving, organizing, and exporting results.⁷ Luther and Kelly and Hoseth provide a similar list of end-user features but also include institutional considerations such as library goals, cost, vendor support, and compatibility with existing technologies.⁸

While these works are helpful for getting a better sense of what to look for when shopping for a web-scale discovery service, they do not offer guidance on how to design a structured evaluation plan. Indeed, many library evaluations have tended to rely on what can be described as the checklist method of evaluation. This typically involves creating a checklist of desirable features and then evaluating products on the basis of whether they provide these features. For example, in developing an evaluation process for Rider University, Chickering and Yang compiled a list of sixteen user interface features, examined live product installations, and ranked each product according to the number of features offered.⁹ Brubaker, Leach-Murray, and Parker employed a similar process to select a discovery service for the twenty-three members of the Private Academic Library Network of Indiana (PALNI).¹⁰ These types of evaluations suffer from a number of limitations. First, they tend to rely on vendor marketing materials or reviews of implementations at other institutions rather than local trials and testing. Second, product requirements are typically given equal weight rather than prioritized according to importance. Third, these requirements tend to focus predominantly on user interface features while neglecting equally important back end functionality and institutional considerations. Finally, these evaluations do not always include input or participation from library staff, users, and stakeholders.

The first published work to offer a structured model for evaluating web-scale discovery services was Vaughan's "Investigations into Library Web-Scale Discovery Services."¹¹ Vaughan outlines the evaluation process employed at University of Nevada, Las Vegas (UNLV), which, in addition to developing a checklist of product requirements, also included staff surveys, interviews with early adopters, vendor demonstrations, and coverage analysis. The author also provides several useful appendixes with templates and documents that librarians can use to guide their own evaluation. Vaughan's work also appears in Popp and Dallis' must-read compendium *Planning and Implementing Resource Discovery Tools in Academic Libraries*.¹² This substantial volume presents



forty chapters on planning, implementing, and maintaining web-scale discovery services, including an entire section devoted to evaluation and selection. In it, Vaughan elaborates on the UNLV model and offers useful recommendations for creating an evaluation team, educating library staff, and communicating with vendors.¹³ Metz-Wiseman et al. offer an overview of best practices for selecting a web-scale discovery service on the basis of interviews with librarians from fifteen academic institutions.¹⁴ Freivalds and Lush of Penn State University explain how to select a web-scale discovery service through a Request for Proposal (RFP) process.¹⁵ Bietila and Olson describe a series of tests that were done at the University of Chicago to evaluate the coverage and functionality of different discovery tools.¹⁶ Chapman et al. explain how personas, surveys, and usability testing were used to develop a user-centered evaluation process at University of Michigan.¹⁷

The following article attempts to build on this existing literature, combining the best elements from evaluation methods employed at other institutions as well as the author's own, with the aim of providing a comprehensive, step-by-step guide to evaluating web-scale discovery services rooted in best practices.

BACKGROUND

Rutgers, The State University of New Jersey, is a public research university consisting of thirty-two schools and colleges offering degrees in the liberal arts and sciences as well as programs in professional and continuing education. The university is distributed across three regional campuses serving more than 65,000 students and 24,000 faculty and staff. The Rutgers University Libraries comprise twenty-six libraries and centers with a combined collection of more than 10.5 million print and electronic holdings. The Libraries' collections and services support the curriculum of the university's many degree programs as well as advanced research in all major academic disciplines.

In January 2013, the Libraries appointed a cross-departmental team to research, evaluate, and recommend the selection of a web-scale discovery service. The impetus for this initiative derived from a demonstrated need to improve the user search experience on the basis of data collected over the last several years through ethnographic studies, user surveys, and informal interactions at the reference desk and in the classroom. Users reported high levels of dissatisfaction with existing library search tools such as the catalog and electronic databases, which they found confusing and difficult to navigate. Above all, users demanded a simple, intuitive starting point from which to search and access the library's collections. Accordingly, the Libraries began investigating ways to improve access with web-scale discovery. The evaluation team examined offerings from four leading web-scale discovery providers, including EBSCO Discovery Service, ProQuest's Summon, Ex Libris' Primo, and OCLC's WorldCat Local. The process lasted approximately nine months and included extensive product and user research, vendor demonstrations, an RFP, reference interviews, trials, surveys, and product testing. See appendix A for an overview of the evaluation plan.

By the time it began its evaluation, Rutgers was already a latecomer to the discovery game. Most of our peers had already been using web-scale discovery services for many years. However, Rutgers' less-than-stellar experience with federated search had led it to adopt a more cautious attitude toward the latest and greatest of library "holy grails." This wait-and-see approach proved highly beneficial in the end as it allowed time for the discovery market to mature and gave the evaluation team an opportunity to learn from the successes and failures of early adopters. In planning its evaluation, the Rutgers team was able to draw on the experiences of earlier pioneers such as UNLV, Penn State, the University of Chicago, and the University of Michigan. It was on the metaphorical shoulders of these library giants that Rutgers built its own successful evaluation process. What follows is a step-by-step guide for evaluating and selecting a web-scale discovery service on the basis of best practices synthesized from the literature as well as the author's own experiences coordinating the evaluation process at Rutgers. Given the rapidly changing nature of the discovery market, the focus of this article is on the process rather than the results of Rutgers' evaluation. While the results will undoubtedly be outdated by the time this article goes to press, the process is likely to remain relevant and useful for years to come.

Form an Evaluation Team

The first step in selecting a web-scale discovery service is appointing a team that will be responsible for conducting the evaluation. Composition of the team will vary depending on local practice and staffing, but should include representatives from a broad cross section of library units, including collections, public services, technical services, and systems. Institutions with multiple campuses, schools, or library branches will want make sure the interests of these constituencies are also represented. If feasible, the library should consider including actual users on the evaluation team. These may be members of an existing user advisory board or recruits from among the library's student employees and faculty liaisons. Including users on your evaluation team will keep the process focused on user needs and ensure that the library selects the best product to meet them.

There are many reasons for establishing an inclusive evaluation team. First, discovery tools have broad implications for a wide range of library services and functions. Therefore a diversity of library expertise is required for an informed and comprehensive evaluation. Reference and instruction librarians will need to evaluate the functionality of the tool, the quality of results, and its role in the research process. Collections staff will need to assess scope of coverage and congruency with the library's existing subscriptions. Access services will need to assess how the tool handles local holdings information and integrates with borrowing and delivery services like interlibrary loan. Catalogers will need to evaluate metadata requirements and procedures for harvesting local records. IT staff will need to assess technical requirements and compatibility with existing infrastructure and systems.

Second, depending on the size and goals of the institution, the product may be expected to serve a wide community of users with different needs, skill levels, and academic backgrounds. Large



universities that include multiple schools, offer various degree programs, or have specialized programs like law or medicine will need to determine if and how a new discovery tool will address the needs of all these users. It is important that the composition of the evaluation team adequately represents the interests of the different user groups the tool is intended to serve. The evaluation at Rutgers was conducted by a cross-departmental team of fifteen members and included experts from a variety of library units and representatives from all campuses.

Finally, because web-scale discovery brings such profound changes to staff and user workflows, decisions regarding selection and implementation are often fraught with controversy. As noted, discovery tools impact a wide range of library services and therefore require careful evaluation from the perspectives of multiple stakeholders. Furthermore, these tools dramatically change the nature of library research, and not everyone in your organization may view this change as being for the better. Despite growing rates of adoption, debates over the value and utility of web-scale discovery continue to divide librarians.¹⁸ According to one survey, securing staff buy-in is the biggest challenge academic libraries face when implementing a web-scale discovery service.¹⁹ Ensuring broad involvement early in the process will help to secure organizational buy-in and support for the selected product.

While broad representation is important, having a large and diverse team can sometimes slow down the process; schedules can be difficult to coordinate, members may have competing views or demands on their time, meetings can lose focus or wander off topic, etc. The more members on your evaluation team, the more difficult the team may be to manage. One strategy for managing a large group might be to create a smaller, core team with all other members serving on an ad hoc basis. The core team functions as a steering committee to manage the project and calls on the ad hoc members at different stages in the evaluation process where their input and expertise is needed. Another strategy would be to break the larger group into several functional teams, each responsible for evaluating specific aspects of the discovery tool. For example, one team might focus on functionality, another on technology, a third on administration, etc. This method also has the advantage of distributing the workload among team members and breaking down a complex evaluation process into discrete, more manageable parts.

Like any other committee or taskforce, your evaluation team should have a charge outlining its responsibilities, timetable of deliverables, reporting structure, and membership. The charge should also include a vision or goals statement that explicitly states the underlying assumptions and premises of the discovery tool, its purpose, and how it supports the library's larger mission of connecting users with information.²⁰ Although frequently highlighted in the literature, the importance of defining institutional goals for discovery is often overlooked or taken for granted.²¹ Having a vision statement is crucial to the success of the project for multiple reasons. First, it frames the evaluation process by establishing mutually agreed-upon goals and priorities for the product. Before the evaluation can begin, the team must have a clear understanding of what problems the discovery service is expected to solve, who it is intended to serve, and how it

supports the library's strategic goals. Is the service primarily intended for undergraduates, or is it also expected to serve graduate students and faculty? Is it a one-stop shop for all information needs, a starting point in a multi-step research process, or merely a useful tool for general and interdisciplinary research? Second, having a clear vision for the product will help guide implementation and assessment. It will not only help the library decide how to configure the product and what features to prioritize, but also offer explicit benchmarks by which to evaluate performance. Finally, aligning web-scale discovery with the library's strategic plan will help put the project in wider context and secure buy-in across all units in the organization. Having a clear understanding of how the product will be integrated with and support other library services will help minimize common misunderstandings and ensure wider adoption.

Educate Library Stakeholders

Despite the quick maturation and adoption of web-scale discovery services, these technologies are still relatively new. Many librarians in your organization, including those on the evaluation team, may only possess a cursory understanding of what these tools are and how they function. Creating an inclusive evaluation process requires having an informed staff that can participate in the discussions and decision-making processes leading to product selection. Therefore the first task of your evaluation team should be to educate themselves and their colleagues on the ins and outs of web-scale discovery services. This should include performing a literature review, collecting information about products currently on the market, and reviewing live implementations at other institutions.

At Rutgers, the evaluation team conducted an extensive literature review that resulted in annotated bibliography covering all aspects of web-scale discovery, including general introductions, product reviews, and methodologies for evaluation, implementation, and assessment. All team members were encouraged to read this literature to familiarize themselves with relevant terminology, products, and best practices. The team also collected product information from vendor websites and reviewed live implementations at other institutions. In this way, members were able to familiarize themselves with the different features and functionality offered by each vendor.

Once the team has done its research, it can begin sharing its findings with the rest of the library community. Vaughan recommends establishing a quick and easy means of disseminating information such as an internal staff website, blog, or wiki that staff can visit on their own time.²² The Rutgers team created a private LibGuide that served as a central repository for all information related to the evaluation process, including a brief introduction to web-scale discovery, information about each product, recorded vendor demonstrations, links to live implementations, and an annotated bibliography. Also included was information about the team's ongoing work, including the group's charge, timeline, meeting minutes, and reports. In addition to maintaining an online presence, the team also held a series of public forums and workshops to educate staff about the nature of web-scale discovery as well as provide updates on the evaluation process and



respond to questions and concerns. By providing staff with a foundation for understanding web-scale discovery and the process by which these products were to be evaluated, the team sought to maximize the engagement and participation of the larger library community.

Schedule Vendor Demonstrations

Once everyone has a conceptual understanding of what web-scale discovery services do and how they work, it is time to begin inviting onsite vendor demonstrations. These presentations give library staff an opportunity to see these products in action and ask vendors in-depth questions. Sessions are usually led by a sales representative and product manager and typically include a brief history of the product's development, a demonstration of key features and functionality, and an audience question-and-answer period. To provide a level playing field for comparison, the evaluation team may wish to submit a list of topics or questions for each vendor to address in their presentation. This could be a general outline of key areas of interest identified by the evaluation team or a list of specific questions solicited from the wider library community. Vaughan offers a useful list of questions that librarians may wish to consider to structure vendor demonstrations.²³ One tactic used by the evaluation team at Auburn University involved requiring vendors to use their products to answer a series of actual reference questions.²⁴ This not only precluded them from using canned searches that might only showcase the strengths of their products, but also gave librarians a better sense of how these products would perform out in the wild against real user queries. Another approach might be to invite actual users to the demonstrations. Whether you are fortunate enough to have users on your evaluation team or able to encourage a few library student workers to attend, your users may raise important questions that your staff has overlooked.

Vendor demonstrations should only be scheduled after the evaluation team has had an opportunity to educate the wider library community. An informed staff will get more out of the demos and be better equipped to ask focused questions. As Vaughan suggests, demonstrations should be scheduled in close proximity (preferably within the same month) to sustain staff engagement, facilitate retention of details, and make it easier to compare services.²⁵ With the vendor's permission, libraries should also consider recording these sessions and making them available to staff members who are unable to attend. At the conclusion of each demonstration, staff should be invited to offer their feedback on the presentation or ask any follow-up questions. This can be accomplished by distributing a brief paper or online survey to the attendees.

Create an Evaluation Rubric

Perhaps the most important part of the evaluation process is developing a list of key criteria that will be used to evaluate and compare vendor offerings. Once the evaluation team has a better understanding of what these products can do and the different features and functionality offered by each vendor, it can begin defining the ideal discovery environment for its institution. This often takes the form of a list of desirable features or product requirements. The process for generating

these criteria tends to vary by institution. In some cases, they are defined by the team leader or based on criteria used for past technology purchases.²⁶ In other cases, criteria are compiled through a review of the literature.²⁷ In yet other cases, they are developed and refined with input from library staff through staff surveys and meetings.²⁸

One important element missing from all of these approaches is the user. To ensure the evaluation team selects the best tool for library users, product requirements should be firmly rooted in an assessment of user needs. The University of Michigan, for example, used persona analysis to identify common user needs and distilled these into a list of tangible features that could be used for product evaluation.²⁹ Other tactics for assessing user needs and expectations might include user surveys, interviews, or focus groups. These tools can be useful for gathering information about what users want from your web-scale discovery system. However, these methods should be used with caution, as users themselves don't always know what they want, particularly from a product they have never used. Furthermore, as usability experts have pointed out, what users say they *want* may not be what they actually *need*.³⁰ Therefore it is important to validate data collected from surveys and focus groups with usability testing. To reliably determine whether a product meets the needs of your users, it is best to observe what users actually do rather than what they say they do.

If the evaluation team has a short timeframe or is unable to undertake extensive user research, it may be able to develop product requirements on the basis of existing research. At Rutgers, for example, the Libraries' department of planning and assessment conducts a standing survey to collect information about users' opinions of and satisfaction with library services. The evaluation team was able to use this data to learn more about what users like and don't like about the library's current search environment. The team analyzed more than 700 user comments collected from 2009 to 2012 related to the library's catalog and electronic resources. Comments were mapped to specific types of features and functionality that users want or expect from a library search tool. Since most users don't typically articulate their needs in terms of concrete technical requirements, some interpretation was required on the part of the evaluation team. For example, the average user may not necessarily know what faceted browsing is, but a suggestion that there be "a way to browse through books by category instead of always having to use the search box" could reasonably be interpreted as a request for this feature. Features were ranked in order of importance by the number of comments made about it. Some of the most "requested" features included single point of access, "smart" search functionality such as autocorrect and autocomplete, and improved relevance ranking.

Of course, user needs are not the only criteria to be considered when choosing a discovery service. Organizational and staff needs must also be taken into account. User input is important for defining the functionality of the public interface, but staff input is necessary for determining back-end functionality and organizational fit. To the list of user requirements, the evaluation team added institutional requirements related to factors such as cost, coverage, customizability, and

support. The team then conducted a library-wide survey inviting all staff to rank these requirements in order of importance and offer any additional requirements that should be factored into the evaluation.

Combining the input from library staff and users, the evaluation team drafted a list of fifty-five product requirements (see appendix B), which became the basis for a comprehensive evaluation rubric that would be used to evaluate and ultimately select a web-scale discovery service. The design of the rubric was largely modeled after the one developed at Penn State.³¹ Requirements were arranged into five categories: content, functionality, usability, administration, and technology. Each category was allocated to a sub team according to area of expertise that would be responsible for that portion of the evaluation. Each requirement was assigned a weight according to its degree of importance: 3 = mandatory, 2 = desired, 1 = optional. Each product was given a score based on how well it met each requirement: 3 = fully meets, 2 = partially meets, 1 = barely meets, 0 = does not meet. The total number of points awarded for each requirement was calculated by multiplying weight by score. The final score for each product was calculated by summing up the total number of points awarded (see appendix C).

This scoring method was particularly helpful in minimizing the influence of bias on the evaluation process. Keep in mind that some stakeholders may possess personal preferences for or against a particular product because of current or past relations with the vendor, their experiences with the product while at another institution, or their perception of how the product might impact their own work. By establishing a set of predefined criteria, rooted in local needs and measured according to clear and consistent standards, the team adopted an evaluation model that was not only user-centered, but also allowed for a fair, unbiased, and systematic evaluation of vendor offerings. This is particularly important for libraries that must go through a formal procurement process to purchase a web-scale discovery service.

Draft the RFP

Once the evaluation team has defined its product requirements and established a method for evaluating the products in the marketplace, it can set to work drafting a formal RFP. Some institutions may be able to forego the RFP process. Others, like Rutgers, are required to go through a competitive bidding process for any goods and services purchased over a certain dollar amount. The only published model on selecting a discovery service through the RFP process is offered by Freivalds and Lush.³² The authors provide a brief overview of the pros and cons of using an RFP, describe the process developed at Penn State, and offer several useful templates to help guide the evaluation.

The RFP lets vendors know that the organization is interested in their product, outlines the organization's requirements for said product, and gives the vendors an opportunity to explain in detail how their product meets these requirements. RFPs are usually written in collaboration with

your university's purchasing department who typically provides a template for this purpose. At a minimum, your RFP should include the following:

- background information about the library, including size, user population, holdings, and existing technical infrastructure
- a description of the product being sought, including product requirements, services and support expected from the vendor, and the anticipated timeline for implementation
- a summary of the criteria that will be used to evaluate proposals, the deadline for submission, and the preferred format of responses
- any additional terms or conditions such as requiring vendors to provide references, onsite demonstrations, trial subscriptions, or access to support and technical documentation
- information about who to contact regarding questions related to the RFP

RFPs are useful not only because they force the library to clearly articulate its needs for web-scale discovery, but also because they produce a detailed, written record of product information that can be referenced throughout the evaluation process. The key component of Rutgers' RFP was a comprehensive, 135-item questionnaire that asked vendors to spell out in painstaking detail the design, technical, and functional specifications of their products (see appendix D). Many of the questions were either borrowed from the existing literature or submitted by members of the evaluation team. All questions were directly mapped to criteria from the team's evaluation rubric. The responses were used to determine how well each product met these criteria and factored into product scoring. Vendors were given one month to respond to the RFP.

Interview Current Customers

While vendor marketing materials, demonstrations, and questionnaires are important sources of product information, vendor claims should not simply be taken at face value. To obtain an impartial assessment of the products under consideration, the evaluation team should reach out to current customers. There are several ways to identify current discovery service subscribers. Many published overviews of web-scale discovery services offer lists of example implementations for each major discovery provider.³³ Most vendors also provide a list of subscribers on their website or community wiki (or will provide one on request). And, of course, there is also Marshall Breeding's invaluable website, Library Technology Guides, which provides up-to-date information about technology products used by libraries around the world.³⁴ The advanced search allows you to filter libraries by criteria such as type, collection size, geographic area, and ILS, thereby making it easier to identify institutions similar to your own.

As part of the RFP process, all four vendors were required to provide references for three current academic library customers of equivalent size and classification to Rutgers. These twelve references were then invited to take an online survey asking them to share their opinions of and experiences with the product (see appendix E). The survey consisted of a series of Likert-scale questions asking each reference to rate their satisfaction with various functions and features of

their discovery service. This was followed by many in-depth written response questions regarding topics such as coverage, quality of results, interface usability, customization, and support. Follow-up phone interviews were conducted in cases where additional information or clarification was needed.

The surveys permitted the evaluation team to collect feedback from current customers in a way that was minimally obtrusive while allowing for easy analysis and comparison of responses. It also provided a necessary counterbalance to vendor claims by giving the team a much more candid view of each product's strengths and weaknesses. The reference interviews helped highlight issues and areas of concern that were frequently minimized or glossed over in communications with vendors such as gaps in coverage, inconsistent metadata, duplicate results, discoverability of local collections, and problems with known-item searching.

Configure and Test Local Trials

Although the evaluation team should strive to collect as much product information from as many sources as possible, no amount of research can effectively substitute for a good old-fashioned trial evaluation. Conducting trials using the library's own collections and local settings is the best way to gain first-hand insight into how a discovery service works. For some libraries, the expenditure of time and effort involved in configuring a web-scale discovery service can make the prospect of conducting trials prohibitive. As a result, many discovery evaluations tend to rely on testing existing implementations at other institutions. However, this method of evaluation only scratches the surface. For one thing, the evaluation team is only able to observe the front-end functionality of the public interface. But setting up a local trial gives the library an opportunity to peak under the hood and learn about back-end administration, explore configuration and customization options, attain a deeper understanding of the composition of the central index, and get a better feel for what it is like working with the vendor. Second, discovery services are highly customizable and the availability of certain features, functionality, and types of content varies by institution. As Hoepfner points out, no individual site is capable of demonstrating the "full range of possibilities" available from any vendor.³⁵ The presence or absence of certain features has as much to do with local library decisions as they do with any inherent limitations of the product. Finally, establishing trials gives the evaluation team an opportunity to see how a particular discovery service performs within its own local environment. The ability to see how the product works with the library's own records, ILS, link resolver, and authentication system allows the team to evaluate the compatibility of the discovery service with the library's existing technical infrastructure.

At Rutgers, one of the goals of the RFP was to help narrow the pool of potential candidates from four to two. The evaluation team was asked to review vendor responses and apply the evaluation rubric to assign each a preliminary score on the basis of how well they met the library's requirements. The two top-scoring candidates would then be selected for a trial evaluation that would allow the team to conduct further testing and make a final recommendation. However, after the proposals were reviewed, the scores for three of the products were so close that the team

decided to trial all three. The one remaining product scored notably lower than its competitors and was dropped from further consideration.

Configuring trials for three different web-scale discovery services was no easy task, to be sure. An implementation team was formed to work with the vendors to get the trials up and running. The team received basic training for each product and was given full access to support and technical documentation. Working with the vendors, the implementation team set to work loading the library's records and configuring local settings. For the most part, the trials were basic out-of-the-box implementations with minimal customization. The vendors were willing to do much of the configuration work for us, but it was important that the team learn and understand the administrative functionality of each product, as this was an integral part of the evaluation process. All vendors agreed to a three-month trial period during which the evaluation team ran their products through a series of tests assessing three key areas: coverage, usability, and relevance ranking.

The importance of product testing cannot be overstated. As previously mentioned, web-scale discovery affect a wide variety of library services and, in most cases, will likely serve as the central point of access to the library's collections. Before committing to a product, the library should have an opportunity to conduct independent testing to validate vendor claims and ensure that their products function according to the library's expectations. To ensure that critical issues are uncovered, testing should strive to simulate as much as possible the environment and behavior of your users by employing sample searches and strategies that they themselves would use. In fact, wherever possible, users should be invited to participate in testing and offer their feedback about the products under consideration. Testing checklists and scripts must also be created to guide testers and ensure consistency throughout the process. As Mandernach and Condit Fagan point out, although product testing is time-consuming and labor-intensive, it will ultimately save the time of your users and staff who would otherwise be the first to encounter any bugs and help avoid early unfavorable impressions of the product.³⁶

The first test the evaluation team conducted aimed at evaluating the coverage and quality of indexing of each discovery product (see appendix F). Loosely borrowing from methods employed at University of Chicago, twelve library subject specialists were recruited to help assess coverage within their discipline.³⁷ Each subject specialist was asked to perform three search queries representing popular research topics in their discipline and compare the results from each discovery service with respect to breadth of coverage and quality of indexing. In scoring each product, subject specialists were asked to consider the following questions:

- Do the search results demonstrate broad coverage of the variety of subjects, formats, and content types represented in the library's collection?
- Do any particular types of content seem to dominate the results (books, journal articles, newspapers, book reviews, reference materials, etc.)?

-
- Are the library's local collections adequately represented in the results?
 - Do any relevant resources appear to be missing from the search results (i.e., results from an especially relevant database or journal)?
 - Do item records contain complete and accurate source information?
 - Do item records contain sufficient metadata (citation, subject headings, abstracts, etc.) to help users identify and evaluate results?

Participants were asked to rate the performance of each discovery service in terms of coverage and indexing on a scale of 1 to 3 (1 = poor, 2 = average, 3 = good). Although results varied by discipline, one product received the highest average scores in both areas. In their observations, participants frequently noted that it appeared to have better coverage and produce a greater variety of sources while results from the other two products tended to be dominated by specific source types like newspapers or reference books. The same product was also noted to have more complete metadata while the other two frequently produced results that lacked additional information like abstracts and subject terms.

The second test aimed to evaluate the usability of each discovery service. Five undergraduate students of varying grade levels and areas of study were invited to participate in a task-based usability test (see appendix G). The purpose of the test was to assess users' ability to use these products to complete common research tasks and determine which product best meet their needs. Students were asked to use all three products to complete five tasks while sharing their thoughts aloud. For the purposes of testing, products were referred to by letters (A, B, C) rather than name. Because participants were asked to complete the same tasks using each product, it was assumed that their ability to complete tasks might improve as the test progressed. Accordingly, product order was randomized to minimize potential bias. Each session lasted approximately forty-five minutes and included a pre-test questionnaire to collect background information about the participant as well as a post-test questionnaire to ascertain their opinions on the products being tested. Because users were being asked to test three different products, the number of tasks was kept to a minimum and focused only on basic product functionality. More comprehensive usability testing would be conducted after selection to help guide implementation and improve the selected product.

Using each product, participants were asked to find three relevant sources on a topic, email the results to themselves, and attempt to obtain full text for at least one item. Although the team noted potential problems in users' interaction with all of the products, participants had slightly higher success rates with one product over all others. Furthermore, in the post-test questionnaire, four out of five users stated that they preferred this product to the other two, noting that they found it easier to navigate, obtained more relevant results, and had notably less difficulty accessing full text. A follow-up question asked participants how these products compared with the search tools currently offered by the library. Almost all participants cited disappointing previous experiences

with library databases and the catalog and suggested that a discovery tool might make finding materials easier. However, several users also suggested that none these tools were “perfect.” And, while these discovery services may have the “potential” to improve their library experience, all could use a good deal of improvement, particularly with returning relevant results.

Therefore the evaluation team embarked on a third and final test of its top three discovery candidates, the goal of which was to evaluate relevance ranking. While usability testing is helpful for highlighting problems with the design of an interface, it is not always the best method for assessing the quality of results. In user testing, students frequently retrieved or selected results that were not relevant to the topic. It was not always clear whether this outcome was attributable to a flaw in product design or to the users’ own ability to construct effective search queries and evaluate results. Determining relevance is a subjective process and one that requires a certain level of expertise in the relevant subject area. Therefore, to assess relevance ranking among the competing discovery services, the evaluation team turned once again to its library subject specialists.

Echoing countless other user studies, our testing indicated that most users do not often scroll beyond the first page of results. Therefore a discovery service that harvests content from a wide variety of different sources must have an effective ranking algorithm capable of surfacing the most useful and relevant results. To evaluate relevance ranking, subject specialists were asked to construct a search query related to their area of expertise, perform this search in each discovery tool, and rate the relevancy of the first ten results. Results were recorded in the exact order retrieved and ranked on a scale of 0–3 (0 = not relevant, 1 = somewhat relevant, 2 = relevant, 3 = very relevant).

Two values were used to evaluate the relevance-ranking algorithm of each discovery service. Relevance was assessed by calculating *cumulative gain*, or the sum of all relevance scores. For example, if the first ten results returned by a discovery product received a score of 3 because they were all deemed to be “very relevant,” the product would receive a cumulative gain score of 30. Ranking was assessed by calculating *discounted cumulative gain*, which discounts the relevance score of results on the basis of where they appear in the rankings. Assuming that the relevance of results should decrease with rank, each result after the first was associated with a discount factor of $1/\log_2 i$ (where i = rank). The relevance for each result is multiplied by the discount factor to provide the discount gain. For example, a result with a relevance score of 3 but a rank of 4 is discounted through this process to a relevance score of 1.5. Discounted cumulative gain represents the sum of all discount gain scores.³⁸

Eighteen librarians conducted a total of twenty-six searches. Using a Microsoft Excel worksheet, participants were asked to record their search query, the titles of the first ten results, and the relevance score of each result (see appendix H). Formulas for cumulative gain and discount cumulative gain were embedded in the worksheet so these values were automatically calculated. After all the values were calculated, one product once again had outperformed all others. In the

majority of searches conducted, librarians rated its results as being more relevant than its competitors. However, librarians were quick to point out that they were not entirely satisfied with the results from any of the three products. In their observations, they noted many of the same issues that were raised in previous rounds of testing such as incomplete metadata, duplicate results, and overrepresentation of certain types of content.

At the end of the trial period, the evaluation team once again invited feedback from the library staff. An online library-wide survey was distributed in which staff members were asked to rank each discovery product according to several key requirements drawn from the team's evaluation rubric. Each requirement was accompanied by one or more questions for participants to consider in their evaluation. The final question asked participants to rank the three candidates in order of preference. Links to the trial implementations of all three products were included in the survey. Included in the email announcement was also a link to the team's website where participants could find more information about web-scale discovery. Because participating in the survey required staff to review and interact with all three products, the team estimated that it would take forty-five minutes to an hour to complete (depending on the staff member's familiarity with the products). Given the amount of time and effort required for participation, relevant committees were also encouraged to review the trials and submit their evaluation as a group. Response rate for the survey was much lower than expected, possibly because of the amount of effort involved or because a large number of staff did not feel qualified to comment on certain aspects of the evaluation. However, among the staff members that did respond, one product was rated more highly than all others. Coincidentally, it was also the same product that had received the highest scores in all three rounds of testing.

Make Final Recommendation

At this stage in the process, your evaluation team should have collected enough data to make an informed selection decision. Your decision should take into consideration all of the information gathered throughout the evaluation process, including user and product research, vendor demonstrations, RFP responses, customer references, staff and user feedback, trials, and product testing. In preparation for the evaluation team's final meeting, each sub team was asked to revisit the evaluation rubric. Using all of the information that had been collected and made available on the team's website, each sub team was asked to score the remaining three candidates based on how well they met the requirements in their assigned category and submit a report explaining the rationale for their scores. At the final meeting, a representative from each sub team presented their report to the larger group. The entire team reviewed the scores awarded to each product. Once a consensus was reached on the scoring, the final results were tabulated and the product that received the highest total score was selected.

Once the evaluation team has reached a conclusion, its decision needs to be communicated to library stakeholders. The team's findings should be compiled in a final report that includes a brief introduction to the subject of web-scale discovery, the factors motivating the library's decision to

acquire a discovery service, an overview of the methods that were used evaluate these services, and a summary of the team's final recommendation. Of course, considering that few people in your organization may ever actually read the report, the team should seek out additional opportunities to present its findings to the community. The Rutgers evaluation team presented its recommendation report on three different occasions. The first was joint meeting of the library's two major governing councils. After securing the support of the councils, the group's recommendation was presented at a meeting of library administrators for final approval. Once approved, a third and final presentation was given at an all-staff meeting and included a demonstration of the selected product. By taking special care to openly communicate the team's decision and making transparent the process used to reach it, the evaluation team not only demonstrated the depth of its research but also was able to secure organizational buy-in and support for its recommendation.

CONCLUSION

Selecting a web-scale discovery service is a large and important undertaking that involves a significant investment of time, staff, and resources. Finding the right match begins with a thorough and carefully planned evaluation process. The evaluation process outlined here is intended as a blueprint that similar institutions may wish to follow. However, every library has different needs, means, and goals. While this process served Rutgers well, certain elements may not be applicable to your institution. Regardless of what method your library chooses, it should strive to create an evaluation process that is inclusive, goal-oriented, data-driven, user-centered, and transparent.

Inclusive

Web-scale discovery impacts a wide variety of library services and functions. Therefore a complete and informed evaluation requires the participation and expertise of a broad cross section of library units. Furthermore, as with the adoption of any new technology, the implementation of a web-scale discovery service can be potentially disruptive. These products introduce significant and sometimes controversial changes to staff workflows, user behavior, and library usage. Ensuring broad involvement in the evaluation process can help allay potential concerns, reduce tensions, and ensure wider adoption.

Goal-Oriented

It can be easy to be seduced by new technologies simply because they are new. But merely adopting these technologies without taking to the time to reflect on and communicate their purpose and goals can be a recipe for disaster. To select the best discovery tool for your library, evaluators must have a clear understanding of the problems it is trying to solve, the audience it seeks to serve, and the role it plays within the library's larger mission. Articulating the library's vision and goals for web-scale discovery is crucial for establishing an evaluation plan, developing a prioritized list of product requirements, understanding what questions to ask vendors, and setting benchmarks by which to evaluate performance.



Data-Driven

To ensure an informed, fair, and impartial evaluation, evaluators should strive to incorporate data-driven practices into all of their decision-making. Many library stakeholders, including members of the evaluation team, may enter the evaluation process with preexisting views on web-scale discovery, untested assumptions about user behavior, or strong opinions about specific products and vendors. To minimize the influence of these potential biases on the selection process, it is important that the team be able to demonstrate the rationale for its decisions through verifiable data. Evaluating web-scale discovery services requires extensive research and should include data collected through user research, staff surveys, collections analysis, and product testing. All of this data should be carefully collected, analyzed, and used to inform the team's final recommendation.

User-Centered

If the purpose of adopting a web-scale discovery service is to better serve your users, then you should try as much as possible to involve users in the evaluation and selection process. This means including users on the evaluation team, grounding product requirements in user research, and gathering user feedback through surveys, focus groups, and product testing. This last step is especially important. No other piece of information gathered throughout the evaluation process will be as helpful or revealing as actually watching users use these products to complete real-life research tasks. User testing is the best and, frankly, only way to validate claims from both vendors and librarians about what your users want and need from your library's search environment.

Transparent

Because web-scale discovery impacts library staff and users in significant ways, its reception within academic libraries has been somewhat mixed. As previously mentioned, securing staff buy-in is often one of the most difficult obstacles libraries face when introducing a new web-scale discovery service. While encouraging broad participation in the evaluation process helps facilitate buy-in, not every library stakeholder will be able to participate. Therefore it is important that the evaluation team make special effort to communicate its work and keep the library community updated on its progress. This can be done by creating a staff website or blog devoted to the evaluation process, sending periodic updates via the library's electronic discussion list, holding public forums and demonstrations, regularly soliciting staff feedback through surveys and polls, and widely distributing the team's findings and final report. These communications should help secure organizational support by making clear that the team recommendations are based on a thorough evaluation that is inclusive, goal-oriented, data-driven, user-centered, and transparent.

Appendix A. Overview of Web-Scale Discovery Evaluation Plan

1

Form an evaluation team

Create an evaluation team representing a broad cross section of library units. Draft a charge outlining the library's goals for web-scale discovery and the team's responsibilities, timetable, reporting structure, and membership.

2

Educate library stakeholders

Create a staff website or blog to disseminate information about web-scale discovery and the evaluation process. Host workshops and public forums to educate staff, share information, and maximize community participation.

3

Schedule vendor demonstrations

Invite vendors for onsite product demonstrations. Schedule visits in close proximity and provide vendors with an outline or list of questions in advance. Invite all members of the library community to attend and offer feedback.

4

Create an evaluation rubric

Create a comprehensive, prioritized list of product requirements rooted in staff and user needs. Develop a fair and consistent scoring method for determining how each product meets these requirements.

5

Draft the RFP

If required, draft an RFP to solicit bids from vendors. Include information about your library, a summary of your product requirements and evaluation criteria, and any terms or conditions of the bidding process.

6

Interview current customers

Obtain candid assessments of each product by interviewing current customers. Ask customers to share their experiences and offer assessments on factors such as coverage, design, functionality, customizability, and vendor support.

7

Configure and test local trials

After narrowing down the options, select the top candidates for a trial evaluation. Test the products with users and staff to evaluate and compare coverage, functionality, and result quality.

8

Make final recommendation

Make an informed recommendation based on all of the information collected. Compile the results of your research in a final report and communicate the team's findings to the library community.

Appendix B. Product Requirements for a Web-Scale Discovery Service

#	Requirement	Description	Questions to Consider
1	Content		
1.1	Scope	Provides access to the broadest possible spectrum of library content including books, periodicals, audiovisual materials, institutional repository items, digital collections, and open access content	With how many publishers and aggregators does the vendor have license agreements? Are there any notable exclusions? How many total unique items are included in the central index? How many open access resources are included? What percentage of content is mutually licensed? What is the approximate disciplinary, format, and date breakdown of the central index? What types of local content can be ingested into the index (ILS records, institutional repository items, digital collections, research guides, webpages, etc.)? Can the library customize what content is exposed to its users?
1.2	Depth	Provides the richest possible metadata for all indexed items, including citations, descriptors, abstracts, and full text	What level of indexing is provided? What percentage of items contains only citations? What percentage includes abstracts? What percentage includes full text?
1.3	Currency	Provides regular and timely updates of licensed content as well as on-demand updates of local content	How frequently is the central index updated? How frequently are local records ingested? Can the library initiate a manual harvest of local records? Can the library initiate a manual harvest of a specific subset of local records?

1.4	Data quality	Provides clear and consistent indexing of records from a variety of different sources and in a variety of different formats	What record formats are supported? What metadata fields are required for indexing? How is metadata from different sources normalized into a universal metadata schema? How are controlled vocabularies created? To what degree can collections from different sources have their own unique field information displayed and/or calculated into the relevancy-ranking algorithm for retrieval purposes?
1.5	Language	Supports indexing and searching of foreign-language materials using non-Roman characters	Does the product support indexing and searching of foreign-language materials using non-Roman characters? What languages and character sets are supported?
1.6	Federated searching	Supports incorporation of content not included in the central index via federated searching	Does the vendor offer federated searching of sources not included in the central index? How are these sources integrated into search results? Is there an additional cost for adding connectors to these sources?
1.7	Unlicensed content	Includes and makes discoverable additional content not owned or licensed by the library	Are local collections from other libraries using the discovery service exposed to all customers? Are users able to search content that is included in the central index but not licensed or owned by the host library?
2	Functionality		
2.1	Smart searching	Provides “smart” search features such as autocomplete, autocorrect, autostemming, thesaurus matching, stop-word filtering, keyword highlighting, etc.	What “smart” features are included in the search engine? Are these features customizable? Can they be enabled or disabled by the library?

2.2	Advanced searching	Provides advanced search options such as field searching, Boolean operators, proximity searching, nesting, wildcard/truncation, etc.	What types of advanced search options are available? Are these options customizable? Can they be enabled or disabled by the library?
2.3	Search limits	Provides limits for refining search results according to specified criteria such as peer-review status, full-text availability, or location	Does the product include appropriate limits for filtering search results?
2.4	Faceted browsing	Allows users to browse the index by facets such as format, author, subject, region, era, etc.	What types of facets are available for browsing? Can users select multiple facets in different categories? Are facets easy to add or remove from a search? Are facet categories, labels, and ordering customizable? Can facets be customized by format or material type (e.g., music, film, etc.)?
2.5	Scoped searching	Provides discipline-, format-, or location-specific search options that allow searches to be limited to a set of predefined resources or criteria	Can the library construct scoped search portals for specific campus libraries, disciplines, or formats? Can these portals be customized with different search options, facets, relevancy ranking, or record displays?
2.6	Visual searching	Provides visual search and browse options such as tag clouds, cluster maps, virtual shelf browsing, geo-browsing, etc.	Does the product provide any options for visualizing search results beyond text-based lists? Can data visualization tools be integrated into search result display with additional programming?
2.7	Relevancy ranking	Provides useful results using an effective and locally customizable relevancy ranking algorithm	What criteria are used to determine relevancy (term frequency and placement, format, document length, publication date, user behavior, scholarly value, etc.)? How does it rank items with varying levels of metadata (e.g., citation only vs. citation + full text)? Is relevancy ranking customizable

			by the library? By the user?
2.8	Deduplication	Has an effective method for identifying and managing duplicate records within results	Does the product employ an effective method of deduplication?
2.9	Record grouping	Groups different manifestations of the same work together in a single record or cluster	Does the product employ FRBR or some similar method to group multiple manifestations of the same work?
2.10	Result sorting	Provides alternative options for sorting results by criteria such as date, title, author, call number, etc.	What options does the product offer for sorting results?
2.11	Item holdings	Provides real-time local holdings and availability information within search results	How does the product provide local holdings and availability information? Is this information displayed in real-time? Is this information displayed on the results screen or only within the item record?
2.12	OpenURL	Supports openURL linking to facilitate seamless access from search results to electronic full text and related services	How does the product provide access to the library's licensed full-text content? Are openURL links displayed on the results screen or only in the item record?
2.13	Native record linking	Provides direct links to original records in their native source	Does the product offer direct links to original records allowing users to easily navigate from the discovery service to the record source, whether it is a subscription database, the library catalog, or the institutional repository?
2.14	Output options	Provides useful output options such as print, email, text, cite, export, etc.	What output options does the product offer? What citation formats are supported? Which citation managers are supported? Are export options customizable?

2.15 Personalization	Provides personalization features that allow users to customize preferences, save results, bookmark items, create lists, etc.	What personalization features does the product offer? Are these features linked to a personal account or only session-based? Must users create their own accounts or can accounts be automatically linked to their institutional ID?
2.16 Recommendations	Provides recommendations to help users locate similar items or related resources	Does the product provide item recommendations to help users locate similar items? Does the product provide database recommendations to help users identify specialized databases related to their topic?
2.17 Account management	Allows users to access their library account for activities such as renewing loans, placing holds and requests, paying fines, viewing borrowing history, etc.	Can the product be integrated with the library's ILS to provide seamless access to user account management functions? Does the vendor provide any drivers or technical support for this purpose?
2.18 Guest access	Allows users to search and retrieve records without requiring authentication	Does the vendor allow for "guest access" to the service? Are users required to authenticate to search or only when requesting access to licensed content?
2.19 Context-sensitive services	Interacts with university identity and course-management systems to deliver customized services on the basis of user status and affiliation	Can the product be configured to interact with university identity and course-management systems to deliver customized services on the basis of user status and affiliation? Does the vendor provide any drivers or technical support for this purpose?
2.20 Context-sensitive delivery options	Displays context sensitive delivery options based on the item's format, status, and availability	Can the product be configured to interact with the library's ILL and consortium borrowing services to display context-sensitive delivery options for unavailable local holdings? Does the vendor provide any drivers or technical support for this purpose?

2.21 Location mapping	Supports dynamic library mapping to help users physically locate items on the shelf	Can the product be configured to support location mapping by linking the call numbers of physical items to online library maps? What additional programming is required?
2.22 Custom widgets	Supports the integration of custom library widgets such as live chat	Can the library's chat service be embedded into the interface to provide live user support? Where can it be embedded? Search page? Result screen?
2.23 Featured items	Highlights new, featured, or popular items such as recent acquisitions, recreational reading, or heavily borrowed or downloaded items	Can the product be configured to dynamically highlight specific items or collections in the library?
2.24 Alerts	Provides customizable alerts or RSS feeds to inform users about new items related to their research or area of study	Does the product offer customizable alerts or RSS feeds?
2.25 User-submitted content	Supports user-submitted content such as tags, ratings, comments, and reviews	What types of user-submitted content does the product support? Is this content only available to the host library or is it shared among all subscribers of the service? Can these features be optionally enabled or disabled?
2.26 Social media integration	Allows users to seamlessly share items via social media such as Facebook, Twitter, Delicious, etc.	What types of social media sharing does the product support? Can these features be enabled or disabled?



3	Usability		
3.1	Design	Provides a modern, aesthetically appealing design that is locally customizable	Does the product have a modern, aesthetically pleasing design? Is it easy to locate all important elements of the interface? Are colors, graphics, and spacing used effectively to organize content? What aspects of the interface are locally customizable (color scheme, branding, navigation menus, result display, item records, etc.)? Can the library apply its own custom stylesheets or is customization limited to a set or predefined options?
3.2	Navigation	Provides an interface that is easy to use and navigate with little or no specialized knowledge	Is the interface intuitive and easy to navigate? Does it use familiar navigational elements and intuitive icons and labels? Are links clearly and consistently labeled? Do they allow the user to easily move from page to page (forward and back)? Do they take the user where he or she expects to go?
3.3	Accessibility	Meets ADA and Section 508 accessibility requirements	Does the product meet ADA and Section 508 accessibility requirements?
3.4	Internationalization	Provides translations of the user interface in multiple languages	Does the vendor offer translations of the interface in multiple languages? Which languages are supported? Does this include translations of customized text?
3.5	Help	Provides user help screens that are thorough, easy to understand, context-sensitive, and customizable	Are product help screens thorough, easy to navigate, and easy to understand? Are help screens general or context-sensitive (i.e., relevant to the user's current location within the system)? Are help screens customizable?

3.6	Record display	Provides multiple record displays with varying levels of information (e.g., preview, brief view, full view, staff view, etc.)	Are record displays well organized and easily scannable? Does the product offer multiple record displays with varying levels of information? What types of record displays are available? Can record displays be customized by item type or search portal?
3.7	Enriched content	Supports integration of enriched content from third-party providers such as cover images, table of contents, author biographies, reviews, excerpts, journal rankings, citation counts, etc.	What types of enriched content does the vendor provide or support? Is there an additional cost for this content?
3.8	Format icons	Provides intuitive icons to indicate the format of items within search results	Does the product provide any icons or visual cues to help users easily recognize the formats of the variety of items displayed in search results? Is this information displayed on the results screen or only within the item record? How does the product define formats? Are these definitions customizable?
3.9	Persistent URLs	Provides short, persistent links to item records, search queries, and browse categories	Does the product offer persistent links to item records? What about persistent links to canned searches and browse categories? Are these links sufficiently short and user-friendly?
4	Administration		
4.1	Cost	Is offered at a price that is within the library's budget and proportional to the value of the service	How is product pricing calculated? What is the total cost of the service including initial upfront costs and ongoing costs for subscription and technical support? What additional costs would be incurred for add-on services (e.g., federated search, recommender services, enriched content, customer support, etc.)?
4.2	Implementation	Is capable of being implemented within the	What is the estimated timeframe for implementation, including



		library's designated timeframe	loading of local records and configuration and customization of the platform?
4.3	User community	Is widely used and respected among the library's peer institutions	How many subscribers does the product have? What percentage of subscribers are college or university libraries? How do current subscribers view the service?
4.4	Support	Is supported by high-quality customer service, training, and product documentation	Does the vendor provide adequate support, training, and help documentation? What forms of customer support are offered? How adequate is the vendor's documentation regarding content agreements, metadata schema, ranking algorithms, APIs, etc.? Does the vendor provide on-site and online training? Is there any additional cost associated with training?
4.5	Administrative tools	Is supported by a robust, easy-to-use administrative interface and customization tools	Does the product have an easy to use administrative interface? Does it support multiple administrator logins and roles? What tools are provided for product customization and administering access control?
4.6	Statistics reporting	Includes a robust statistical reporting modules for monitoring and analyzing product usage	Does the vendor offer a means of capturing and reporting system and usage statistics? What kinds of data are included in such reports? In what formats are these reports available? Is the data exportable?

5	Technology		
5.1	Development	Is a sufficiently mature product supported by a stable codebase and progressive development cycle	Is the product sufficiently mature and supported by a stable codebase? Is development informed by a dedicated user's advisory group? How frequently are improvements and enhancements made to the service? Is there a formal mechanism by which customers can suggest, rank, and monitor the status of enhancement requests? What major enhancements are planned for the next 3–5 years?
5.2	Authentication	Is compatible with the library's authentication protocols	Does the product allow for IP-authentication for on-site users and proxy access for remote users? What authentication methods are supported (e.g., LDAP, CAS, Shibboleth, etc.)?
5.3	Browser compatibility	Is compatible with all major web browsers	What browsers does the vendor currently support?
5.4	Mobile access	Is accessible on mobile devices	Is the product accessible on mobile devices via a mobile optimized web interface or app? Does the mobile version include the same features and functionality of the desktop version?
5.5	Portability	Can be embedded in external platforms such as library research guides, course management systems, or university portals	Can custom search boxes be created and embedded in external platforms such as library research guides, course management systems, or university portals?

5.6	Interoperability	Includes a robust API and is interoperable with other major library systems such as the ILS, ILL, proxy server, link resolver, institutional repository, etc.	Is the product interoperable with other major library systems such as the ILS, ILL, proxy server, link resolver, institutional repository, etc.? Does the vendor offer a robust API that can be used to extract data from the central index or pair it with a different interface? What types of data can be extracted with the API?
5.7	Consortia support	Supports multiple product instances or configurations for a multilibrary environment	Can the technology support multiple institutions on the same installation, each with its own unique instance and configuration of the product? Is there an additional cost for this service?

Appendix C. Sample Web-Scale Discovery Evaluation Rubric

Category	Functionality			
Product	Product A			
Requirement	Weight	Score	Points	Notes
2.1 Smart searching				
2.2 Advanced searching	Weight Scale 1 = Optional 2 = Desired 3 = Mandatory		$Points = Weight \times Score$	Explanation and rationale for score
2.3 Search limits				
2.4 Faceted browsing				
2.5 Scoped searching				
2.6 Visual searching		Scoring Scale 0 = Does not meet 1 = Barely meets 2 = Partially meets 3 = Fully meets		
2.7 Relevancy ranking				
2.8 Deduplication				
2.9 Record grouping				
2.10 Result sorting				
2.11 Item holdings				
2.12 OpenURL				
2.13 Native record linking				
2.14 Output options				
2.11 Item holdings				

Appendix D. Web-Scale Discovery Vendor Questionnaire

1. Content

1.1 Scope

With how many content publishers and aggregators have you forged content agreements?

Are there any publishers or aggregators with whom you have exclusive agreements that prohibit or limit them from making their content available to competing discovery vendors? If so, which ones?

Does your central index exclude any of the publishers and aggregators listed in appendix Y [not reproduced here]? If so, which ones?

How many total unique items are included in your central index?

What is the approximate disciplinary breakdown of the central index? What percentage of content pertains to subjects in the humanities? What percentage in the sciences? What percentage in the social sciences?

What is the approximate format breakdown of the central index? What percentage of content derives from scholarly journals? What percentage derives from magazines, newspapers, and trade publications? What percentage derives from conference proceedings? What percentage derives from monographs? What percentage derives from other publications?

What is the publication date range of the central index? What is the bulk publication date range (i.e., the date range in which the majority of content was published)?

Does your index include content from open access repositories such as DOAJ, HathiTrust, and arXiv? If so, which ones?

Does your index include OCLC WorldCat catalog records? If so, do these records include holdings information?

What types of local content can be ingested into the index (e.g., library catalog records, institutional repository items, digital collections, research guides, library web pages, etc.)?

Can your service host or provide access to items within a consortia or shared catalog like the Pennsylvania Academic Library Consortium (PALCI) or Committee on Institutional Cooperation (CIC)?

Are local collections (ILS records, digital collections, institutional repositories, etc.) from libraries that use your discovery service exposed to all customers?

Can the library customize its holdings within the central index? Can the library choose what content to expose to its users?

1.2 Depth

What level of indexing do you typically provide in your central index? What percentage of items contains only citations? What percentage includes abstracts? What percentage includes full text?

1.3 Currency

How frequently is the central index updated?

How often do you harvest and ingest metadata for the library's local content? How long does it typically take for such updates to appear in the central index?

Can the library initiate a manual harvest of local records? Can the library initiate a manual harvest of a specific subset of local records?

1.4 Data quality

With what metadata schemas (MARC, METS, MODS, EAD, etc.) does your discovery platform work?

Do you currently support RDA records? If not, do you have any plans to do so in the near future?

What metadata is required for a local resource to be indexed and discoverable within your platform?

How is metadata from different sources normalized into a universal metadata schema?

To what degree can collections from different sources have their own unique field information displayed and/or calculated into the relevancy-ranking algorithm for retrieval purposes?

Do you provide authority control? How are controlled vocabularies for subjects, names, and titles established?

1.5 Language

Does your product support indexing and searching of foreign language materials using non-Roman characters? What languages and character sets are supported?

1.6 Federated searching

How does your product make provisions for sources not included in your central index? Is it possible to incorporate these sources via federated search? How are federated search results

displayed with the results from the central index? Is there an additional cost for implementing federated search connectors to these resources?

1.7 Unlicensed content

Are end users able to search content that is included in your central index but not licensed or owned by the library? If so, does your system provide a locally customizable message to the user or does the user just receive the publisher/aggregator message encouraging them to purchase the article? Can the library opt not to expose content it does not license to its users?

2. Functionality

2.1 “Smart” searching

Does your product include autocomplete or predictive search functionality? How are autocomplete predictions populated?

Does your product include autocorrect or “did you mean . . .” suggestions to correct misspelled queries? How are autocorrect suggestions populated?

Does your product support search query stemming to automatically retrieve search terms with variant endings (e.g., car/cars)?

Does your product support thesaurus matching to retrieve synonyms and related words (e.g., car/automobile)?

Does your product support stop word filtering to automatically remove common stop words (e.g., a, an, on, from, the, etc.) from search queries?

Does your product support search term highlighting to automatically highlight search terms found within results?

How does your product handle zero result or “dead end” searches? Please describe what happens when a user searches for an item that is not included in the central index or the library’s local holdings but may be available through interlibrary loan.

Does your product include any other “smart” search features that you think enhance the usability of your product?

Are all of the above mentioned search features customizable by the library? Can they be optionally enabled or disabled?

2.2 Advanced searching

Does your product support Boolean searching that allows users to combine search terms using operators such as AND, OR, and NOT?

Does your product support fielded searching that allows users to search for terms within specific metadata fields (e.g., title, author, subject, etc.)?

Does your product support phrase searching that allows users to search for exact phrases?

Does your product support proximity searching that allows users to search for terms within a specified distance from one another?

Does your product support nested searching to allow users to specify relationships between search terms and determine the order in which they will be searched?

Does your product support wildcard and truncation searching that allow users to retrieve variations of their search terms?

Does your product include any other advanced search features that you think enhance the usability of your product?

Are all of the above mentioned search features customizable by the library? Can they be optionally enabled or disabled?

2.3 Search limits

Does your product offer search limits for limiting results according to predetermined criteria such as peer-review status or full text availability?

2.4 Faceted browsing

Does your product support faceted browsing of results by attributes such as format, author, subject, region, era, etc.? If so, what types of facets are available for browsing?

Is faceted browsing possible before as well after the execution of a search?

Can users select multiple facets in different categories?

Are facet categories, labels, and ordering customizable by the library?

Can specialized materials be assigned different facets in accordance with their unique attributes (e.g., allowing users to browse music materials by unique attributes such as medium of performance, musical key/range, recording format, etc.)?

2.5 Scoped searching

Does your product support the construction of multiple scoped search portals for specific campus libraries, disciplines (medicine), or formats (music/video)?

If so, what aspects of these search portals are customizable (branding, search options, facets, relevancy ranking, record displays, etc.)?

2.6 Visual searching

Does your product provide any options for visualizing search results beyond text-based lists, such as cluster maps, tag clouds, image carousels, etc.?

2.7 Relevancy ranking

Please describe your relevancy ranking algorithm. In particular, please describe what criteria are used to determine relevancy (term frequency/placement, item format/length, publication date, user behavior, scholarly value, etc.) and how is each weighted?

How does your product rank items with varying levels of metadata (e.g., citation only vs. citation, abstract, and full text)?

Is relevancy ranking customizable by the library?

Can relevancy ranking be customized by end users?

2.8 Deduplication

How does your product identify and manage duplicate records?

2.9 Record grouping

Does your product employ a FRBR-ized method to group different manifestations of the same work?

2.10 Result sorting

What options does your product offer for sorting results?

2.11 Item holdings

How does your product retrieve and display availability data for local physical holdings? Is there a delay in harvesting this data or is it presented in real time? Is item location and availability displayed in the results list or only in the item record?

2.12 OpenURL

How does your product provide access to the library's licensed full text content?

Are openURL links displayed on the results screen or only in the item record?

2.13 Native record linking

Does your product offer direct links to original records in their native source (e.g., library catalog, institutional repository, third-party databases, etc.)?

2.14 Output options

What output options does your product offer (e.g., print, save, email, SMS, cite, export)?

If you offer a citation function, what citation formats does your product support (MLA, APA, Chicago, etc.)?

If you offer an export function, which citation managers does your product support (e.g., RefWorks, EndNote, Zotero, Mendeley, EasyBib, etc.)?

Are citation and export options locally customizable? Can they be customized by search portal?

2.15 Personalization

Does your product offer any personalization features that allow users to customize preferences, save results, create lists, bookmark items, etc.? Are these features linked to a personal account or are they session-based?

If personal accounts are supported, must users create their own accounts or can account creation be based on the university's CAS/LDAP identity management system?

2.16 Recommendations

Does your product provide item recommendations to help users locate similar items? On what criteria are these recommendations based?

Is your product capable of referring users to specialized databases based on their search query? (For example, can a search for "autism" trigger database recommendations suggesting that the user try their search in PsycINFO or PubMed?) If so, does your product just provide links to these resources or does it allow the user to launch a new search by passing their query to the recommended database?

2.17 Account management

Can your product be integrated with the library's ILS (SirsiDynix Symphony) to provide users access to its account management functions (e.g., renewing loans, placing holds/requests, viewing borrowing history, etc.)? If so, do you provide any drivers or technical support for this purpose?

2.18 Guest access

Are users permitted “guest access” to the service? Are users required to authenticate in order to search or only when requesting access to licensed content?

2.19 Context-sensitive services

Could your product be configured to interact with our university course management systems (Sakai, Blackboard, and eCollege) to deliver customized services based on user status and affiliation? If so, do you provide any drivers or technical support for this purpose?

2.20 Context-sensitive delivery options

Could your product be configured to interact with the library’s interlibrary loan (ILLiad) and consortium borrowing services (EZBorrow and UBorrow) to display context-sensitive delivery options for unavailable local holdings? If so, do you provide any drivers or technical support for this purpose?

2.21 Location mapping

Could your product be configured to support location mapping by linking the call numbers of physical items to library maps?

2.22 Custom widgets

Does your product support the integration of custom library widgets such as live chat? Where can these widgets be embedded?

2.23 Featured items

Could your product be configured to highlight specific library items such as recent acquisitions, popular items, or featured collections?

2.24 Alerts

Does your product offer customizable alerts or RSS feeds to inform users about new items related to their research or area of study?

2.25 User-submitted content

Does your product support user-generated content such as tags, ratings, comments, and reviews?

Is user-generated content only available to the host library or is it shared among all subscribers of your service?

Can these features be optionally enabled or disabled?

2.26 Social media integration

Does your product allow users to seamlessly share items via social media such as Facebook, Google+, and Twitter?

Can these features be optionally enabled or disabled?

3. Usability

3.1 Design

Describe how your product incorporates established best practices in usability. What usability testing have you performed and/or do you conduct on an ongoing basis?

What aspects of the interface's design are locally customizable (e.g., color scheme, branding, display, etc.)?

Can the library apply its own custom stylesheets or is customization limited to a set or predefined options?

3.2 Navigation

What aspects of the interface's navigation are locally customizable (e.g., menus, pagination, facets, etc.)?

3.3 Accessibility

Does your product meet ADA and Section 508 accessibility requirements? What steps have you taken beyond Section 508 requirements to make your product more accessible to people with disabilities?

3.4 Internationalization

Do you offer translations of the interface in multiple languages? Which languages are supported? Does this include translation of any locally customized text?

3.5 Help

Does your product include help screens to assist users in using and navigating the system?

Are help screens general or context-sensitive (i.e., relevant to the user's current location within the system)?

Are help screens locally customizable?

3.6 Record display

Does your product offer multiple record displays with varying levels of information? What types of record displays are available (e.g., preview, brief view, full view, staff view, etc.)?

Can record displays be customizable by item type or metadata (e.g., MARC-based book record vs. MODS-based repository record)?

Can record displays be customizable by search portal (e.g., a biosciences search portal that displays medical rather than LC subject headings and call numbers)?

3.7 Enriched content

Does your product provide or support the integration of enriched content such as cover images, tables of contents, author biographies, reviews, excerpts, journal rankings, citation counts, etc.? If so, what types of content does this include? Is there an additional cost for this content?

3.8 Format icons

Does your product provide any icons or visual cues to help users easily recognize the formats of the variety of items displayed in search results?

How does your product define formats? Are these definitions readily available to end users? Are these definitions customizable?

3.10 Persistent URLs

Does your product offer persistent links to item records?

Does your product offer persistent links to search queries and browse categories?

4. Administration

4.1 Cost

Briefly describe your product pricing model for academic library customers.

4.2 Implementation

Can you meet the timetable defined in appendix Z [not reproduced here]? If not, which milestones cannot be met or which conditions must the Libraries address in order to meet the milestones?

Are you currently working on web-scale discovery implementations at any other large institutions?

4.3 User community

How many live, active installations (i.e., where the product is currently available to end-users) do you currently have?

How many additional customers have committed to the product?

How many of your total customers are college or university libraries?

4.4 Support

What customer support services and hours of availability do you provide for reporting and/or troubleshooting technical problems?

Do you have a help ticket tracking system for monitoring and notifying clients of the status of outstanding support issues?

Do you offer a support website with up-to-date product documentation, manuals, tutorials, and FAQs?

Do you provide on-site and online training for library staff?

Do you provide on-site and online training for end users?

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

Do you have regular public meetings for users to share experiences and provide feedback on the product? If so, where and how often are these meetings held?

What other communication avenues do you provide for users to communicate with your company and also with each other (e.g., listserv, blog, social media)?

4.5 Administration

What kinds of tools are provided for local administration and customization of the product?

Does your product support multiple administrator logins and roles?

4.6 Statistics reporting

What statistics reporting capabilities are included with your product? What kinds of data are available to track and assess collection management and product usage? In what formats are these reports available? Is the data exportable?

Is it possible to integrate third-party analytic tools such as Google Analytics in order to collect usage data?

5. Technology

5.1 Development

In what month and year did product development begin?

What key features differentiate your product from those of your competitors?

How frequently are enhancements and upgrades made to the service?

Please describe the major enhancements you expect to implement in the next year.

Please describe the future direction or major enhancements you envision for the product in the next 3–5 years.

Is there a formal mechanism by which customers may make, rank, and monitor the status of enhancement requests?

Do you have a dedicated user's advisory group to test and provide feedback on product development?

5.2 Authentication

What authentication methods does your product support (e.g., LDAP, CAS, Shibboleth, etc.)?

5.3 Browser compatibility

Please provide a list of currently supported web browsers.

5.4 Mobile access

Is the product accessible on mobile devices via a mobile optimized web interface or app?

Does the mobile version include the same features and functionality of the desktop version?

5.5 Portability

Can custom search boxes be created and embedded in external platforms such as the library's research guides, course management systems, or university portals?

5.6 Interoperability

Does your product include an API that can be used extract data from the central index or pair it with a different interface? What types of data can be extracted with the API? Do you provide documentation and instruction on the functionality and use of your API?

Are there any known compatibility issues with your product and any of the following systems or platforms?

- Drupal
- VuFind
- SirsiDynix Symphony
- Fedora Commons
- EZProxy
- ILLiad

5.7 Consortia support

Can your product support multiple institutions on the same installation, each with its own unique instance and configuration of the product? Is there any additional cost for this service?

Appendix E. Web-Scale Discovery Customer Questionnaire

Institutional Background

Please tell us a little bit about your library.

What is the name of your college or university?

Which web-scale discovery service is currently in use at your library?

- EBSCO Discovery Service (EDS)
- Primo Central (Ex Libris)
- Summon (ProQuest)
- WorldCat Local (OCLC)
- Other _____

When was your current web-scale discovery service selected (month, year)?

How long did it take to implement (even in beta form) your current web-scale discovery service?

Which of the following types of content are included in your web-scale discovery service? (Check all that apply)

- Library catalog records
- Periodical indexes and databases
- Open access content
- Institutional repository records
- Local digital collections (other than your institutional repository)
- Library research guides
- Library web pages
- Other _____

Rate Your Satisfaction

On a scale of 1 (low) to 5 (high), please rate your satisfaction with the following aspects of your web-scale discovery service.

Content

How satisfied are you with the scope, depth, and currency of coverage provided by your web-scale discovery service? **[Are the question marks below the wrong font?]**

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Functionality

How satisfied are you with the search functionality, performance, and result quality of your web-scale discovery service?

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Usability

How satisfied are you with the design, layout, navigability, and overall ease of use of your web-scale discovery interface?

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Administration

How satisfied are you with the administrative, customization, and reporting tools offered by your web-scale discovery service?

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Technology

How satisfied are you with the level of interoperability between your web-scale discovery service and other library systems such as your ILS, knowledge base, link resolver, and institutional repository?

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Overall

Overall, how satisfied are you with your institution's web-scale discovery service?

☺ 1

☺ 2

☺ 3

☺ 4

☺ 5

Questions

Please share your experiences with your web-scale discovery service by responding to the following questions.

Briefly describe your reasons for implementing a web-scale discovery service. What role does this service play at your library? How is it intended to benefit your users? What types of users is it intended to serve?

Does your web-scale discovery service have any notable gaps in coverage? If so, how do you compensate for those gaps or make users aware of resources that are not included in the service?

Are you satisfied with the relevance of the results returned by your web-scale discovery service? Have you noticed any particular anomalies within search results?

Does your web-scale discovery service lack any specific features or functions that you wish were available?

Are there any particular aspects of your web-scale discovery service that you wish were customizable but are not?

Did you face any particular challenges integrating your web-scale discovery service with other library systems such as your ILS, knowledge base, and link resolver?

How responsive has the vendor been in providing technical support, resolving problems, and responding to enhancement requests? Have they provided adequate training and documentation to support your implementation?

In general, how have users responded to the introduction of this service? Has their response been positive, negative, or mixed?

In general, how have librarians responded to the introduction of this service? Has their response been positive, negative, or mixed?

What has been the impact of implementing a web-scale discovery service on the overall usage of your collection? Have you noticed any fluctuations in circulation, full text downloads, or usage of subject-specific databases?

Has your institution conducted any assessment or usability studies of your web-scale discovery service? If so, please briefly describe the key findings of these studies.

Please share any additional thoughts or advice that you think might be helpful to other libraries currently exploring web-scale discovery services.

Appendix F. Sample Worksheet for Web-Scale Discovery Coverage Test

Instructions

Construct 3 search queries representing commonly researched topics in your discipline. Test your queries in each discovery product and compare the results. For each product, record the number of results retrieved and rate the quality of coverage and indexing. Use the space below your ratings to explain your rationale and record any notes or observations.

Rate coverage and indexing a scale of 1 to 3 (1 = POOR, 2 = AVERAGE, 3 = GOOD). In your evaluation, please consider the following:

Coverage

- Do the search results demonstrate broad coverage of the variety of subjects, formats, and content types represented in the library's collection? (Hint: Use facets to examine the breakdown of results by source type or collection).
- Do any particular types of content seem to dominate the results (books, journal articles, newspapers, book reviews, reference materials, etc.)?
- Are the library's local collections adequately represented in the results?
- Do any relevant resources appear to be missing from the search results (e.g., results from an especially relevant database or journal)?

Indexing

- Do item records contain complete and accurate source information?
- Do item records contain sufficient metadata (citation, subject headings, abstracts, etc.) to help users identify and evaluate results?

Example			
Product	Product B		
Reviewer	Reviewer #2		
Discipline	History		
Query	Results	Coverage	Indexing
KW: slavery AND "united states"	181,457	1 (POOR)	3 (GOOD)
		The majority of results appear to be from newspapers and periodicals. Some items designated as "journals" are actually magazines. There are a large number of duplicate records. Some major works on this subject are not represented in the results.	Depth of indexing varies by publication but most include abstracts and subject headings. Some records only include citations, but citations appear to be complete and accurate.

Appendix G. Sample Worksheet for Web-Scale Discovery Usability Test

Pre-Test Questionnaire

Before beginning the test, ask the user for the following information.

Status Undergraduate Graduate Faculty Staff Other

Major/Department _____

What resource do you use most often for scholarly research? _____

On a scale of 1 to 5, how would you rate your ability to find information using library resources?

Low 1 2 3 4 5 *High*

On a scale of 1 to 5, how would you rate your ability to find information using Google or other search engines?

Low 1 2 3 4 5 *High*

Scenarios

Ask the user to complete the following tasks using each product while sharing their thoughts aloud.

1. You are writing a research paper for your communications course. You've recently been discussing how social media sites like Facebook collect and store large amounts of personal data. You decide to write a paper that answers the question: "Are social networking sites a threat to privacy?" Use the search tool to find sources that will help you support your argument.
2. From the first 10 results, select those that you would use to learn more about this topic and email them to yourself. If none of the results seem useful, do not select any.
3. If you were writing a paper on this topic, how satisfied would you be with these results?
 Very dissatisfied Dissatisfied No opinion Satisfied Very satisfied
4. From the first 10 results, attempt to access an item for which full text is available online.

5. Now that you've seen the first 10 results, what would you do next?

- Decide you have enough information and stop
- Continue and review the next set of results
- Revise your search and try again
- Exit and try your search in another library database (which one?)
- Exit and try your search in Google or another search engine
- Other (please explain)

Post-Test Questionnaire

After the user has used all three products, ask them about their experiences.

Based on your experience, please rank the three search tools you've seen in order of preference.

How would you compare these search tools with the search options currently offered by the library?

Appendix H. Sample Worksheet for Web-Scale Discovery Relevance Test

Instructions

Conduct the same search query in each discovery product and rate the relevance of the first 10 results using the scale provided. For each query, record your search condition, terms, and limiters. For each product, record the first 10 results in the exact order they appear, rank the relevance of each result using the relevance scale, and explain the rationale for your score. All calculations will be tabulated automatically.

Relevance Scale

0 = Not relevant	Not at all relevant to the topic, exact duplicate of a previous result, or not enough information in the record or full text to determine relevance
1 = Somewhat relevant	Somewhat relevant but does not address all of concepts or criteria specified in the search query, e.g., addresses only part of the topic, is too broad or narrow in scope, is not in the specified format, etc.
2 = Relevant	Relevant to the topic, but the topic may not be the primary or central subject of the work, or the work is too brief or dated to be useful; a resource that the user <i>might select</i>
3 = Very relevant	Completely relevant; exactly on topic; addresses all concepts and criteria included in the search query; a resource that the user would <i>likely select</i>

Calculations

Cumulative Gain	Measure of overall relevance based on the sum of all relevance scores.
Discount Factor ($1/\log_2 i$)	Penalization of relevance based on ranking. Assuming that relevance decreases with rank, each result after the first is associated with a discount factor based on log factor 2. Discount factor is calculated as $1/\log_2 i$ where i = rank. The discount factor of result #6 is calculated as 1 divided by the logarithm of 6 with base 2, or $1/\log(6,2) = 0.39$.
Discounted Gain	Discounted relevance score based on ranking. Discounted gain is calculated by multiplying a result's relevance score by its discount factor. The discounted gain of a result with a relevance score of 3 and discount factor of 0.39 is 3×0.39 , or 1.17.
Discounted Cumulative Gain	Measure of overall discounted gain based on the sum of all discount gain scores.

Example								
Product	Product C							
Reviewer	Reviewer #3							
Search Condition	Seeking peer reviewed articles about the impact of media violence on children							
Search Terms	"mass media" AND violence AND children							
Limits	Peer reviewed							
Result	Relevance	Notes	C. Gain	Rank	Relevance	1/Logzi	D. Gain	D.C. Gain
Effects of Media Ratings on Children and Adolescents: A Litmus Test of the Forbidden Fruit Effect	0	Research article suggesting that ratings do not influence children's perceptions of films or video games. Not relevant; does not discuss impact of media violence on children.	19	1	0	1.00	0	9.65
Media Violence Associations with the Form and Function of Aggression among Elementary School Children	3	Research article demonstrating a positive association between media violence exposure and levels of physical and relational aggression in grade school students. Very relevant.		2	3	1.00	3	

Harmful Effects of Media on Children and Adolescents	2	Review article discussing the influence of media on negative child behaviors such as violence, substance abuse, and sexual promiscuity. Relevant but does not focus exclusively on media violence.		3	2	0.63	1.26	
The Influence of Media Violence on Children	3	Review article examining opposing views on media violence and its impact on children. Very relevant.		4	3	0.50	1.5	
Remote Control Childhood: Combating the Hazards of Media Culture in Schools	1	Review article discussing the harmful effects of mass media on child behavior and learning as well as strategies educators can use to counteract them. Somewhat relevant but does not focus exclusively on media violence and discussion is limited to the educational context.		5	1	0.43	0.43	
Media Violence, Physical Aggression, and Relational Aggression in School Age Children	3	Research article on the impact of media violence on childhood aggression in relation to different types of aggression, media, and time periods. Very relevant.		6	3	0.39	1.17	
Do You See What I See? Parent and Child Reports of Parental	2	Research article examining the effectiveness of parental monitoring of children's violent media		7	2	0.36	0.72	



Monitoring of Media		consumption. Relevant but focused less on the effects of media violence than strategies for mitigating them.						
Exposure to Media Violence and Young Children with and Without Disabilities: Powerful Opportunities for Family-Professional Partnerships	2	Review article discussing the impact of media violence on children with and without disabilities and recommendations for addressing this through family-professional partnerships. Relevant but slightly more specific than required.		8	2	0.33	0.66	
KITLESİM İLETİSİM ARAÇLARINDA N TELEVİZYONUN 3-6 YAŞ GRUBUNDAKİ ÇOCUKLARIN DAVRANISLAR I ÜZERINE ETKİSİ.	1	Research article demonstrating a positive correlation between media violence exposure and aggressive behavior in grade school students. Seems very relevant but article is in Turkish.		9	1	0.32	0.32	
Sex and Violence: Is Exposure to Media Content Harmful to Children?	2	Review article discussing how exposure to violent or sexually explicit media influences child behavior and what librarians can do about it. Relevant but less than two pages long.		10	2	0.30	0.60	

REFERENCES

1. Judy Luther and Maureen C. Kelly, "The Next Generation of Discovery," *Library Journal* 136, no. 5 (2011): 66.
2. Athena Hoepfner, "The Ins and Outs of Evaluating Web-Scale Discovery Services," *Computers in Libraries* 32, no. 3 (2012): 8.
3. Kate B. Moore and Courtney Greene, "Choosing Discovery: A Literature Review on the Selection and Evaluation of Discovery Layers," *Journal of Web Librarianship* 6, no. 3 (2012): 145–63, <http://dx.doi.org/10.1080/19322909.2012.689602>.
4. Ronda Rowe, "Web-Scale Discovery: A Review of Summon, EBSCO Discovery Service, and WorldCat Local," *Charleston Advisor* 12, no. 1 (2010): 5-10, <http://dx.doi.org/10.5260/chara.12.1.5>; Ronda Rowe, "Encore Synergy, Primo Central," *Charleston Advisor* 12, no. 4 (2011): 11–15, <http://dx.doi.org/10.5260/chara.12.4.11>.
5. Sharon Q. Yang and Kurt Wagner, "Evaluating and Comparing Discovery Tools: How Close Are We towards the Next Generation Catalog?" *Library Hi Tech* 28, no. 4 (2010): 690–709, <http://dx.doi.org/10.1108/07378831011096312>.
6. Jason Vaughan, "Web Scale Discovery Services," *Library Technology Reports* 47, no. 1. (2011): 5–61, <http://dx.doi.org/10.5860/ltr.47n1>.
7. Hoepfner, "The Ins and Outs of Evaluation Web-Scale Discovery Services."
8. Luther and Kelly, "The Next Generation of Discovery"; Amy Hoseth, "Criteria To Consider When Evaluating Web-Based Discovery Tools," in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 90–103, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch006>.
9. F. William Chickering and Sharon Q. Yang, "Evaluation and Comparison of Discovery Tools: An Update," *Information Technology & Libraries* 33, no. 2 (2014): 5–30, <http://dx.doi.org/10.6017/ital.v33i2.3471>.
10. Noah Brubaker, Susan Leach-Murray, and Sherri Parker, "Shapes in the Cloud: Finding the Right Discovery Layer," *Online* 35, no. 2 (2011): 20–26.
11. Jason Vaughan, "Investigations into Library Web-Scale Discovery Services," *Information Technology & Libraries* 31, no. 1 (2012): 32–82, <http://dx.doi.org/10.6017/ital.v31i1.1916>.
12. Mary P. Popp and Diane Dallis, eds., *Planning and Implementing Resource Discovery Tools in Academic Libraries* (Hershey, PA: Information Science Reference, 2012), <http://dx.doi.org/10.4018/978-1-4666-1821-3>.
13. Jason Vaughan, "Evaluating and Selecting a Library Web-Scale Discovery Service," in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane



-
- Dallis (Hershey, PA: Information Science Reference, 2012), 59–76, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch004>.
14. Monica Metz-Wiseman et al., “Best Practices for Selecting the Best Fit,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 77–89, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch005>.
 15. David Freivalds and Binky Lush, “Thinking Inside the Grid: Selecting a Discovery System through the RFP Process,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 104–21, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch007>.
 16. David Bietila and Tod Olson, “Designing an Evaluation Process for Resource Discovery Tools,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 122–36, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch008>.
 17. Suzanne Chapman et al., “Developing a User-Centered Article Discovery Environment,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 194–224, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch012>.
 18. Lynn D. Lampert and Katherine S. Dabbour, “Librarian Perspectives on Teaching Metasearch and Federated Search Technologies,” *Internet Reference Services Quarterly* 12, no.3/4 (2007): 253–78, http://dx.doi.org/10.1300/J136v12n03_02; William Breitbach, “Web-Scale Discovery: A Library of Babel?” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 637–45, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch038>.
 19. Metz-Wiseman et al., “Best Practices for Selecting the Best Fit,” 81.
 20. Meris A. Mandernach and Jody Condit Fagan, “Creating Organizational Buy-In: Overcoming Challenges to a Library-Wide Discovery Tool Implementation,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 422, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch024>.
 21. David P. Brennan, “Details, Details, Details: Issues in Planning for, Implementing, and Using Resource Discovery Tools,” in *Planning and Implementing Resource Discovery Tools in Academic Libraries*, ed. Mary P. Popp and Diane Dallis (Hershey, PA: Information Science Reference, 2012), 44–56, <http://dx.doi.org/10.4018/978-1-4666-1821-3.ch003>; Hoseth, “Criteria To Consider When Evaluating Web-Based Discovery Tools”; Mandernach and Condit Fagan, “Creating Organizational Buy-In.”

-
22. Vaughan, "Evaluating and Selecting a Library Web-Scale Discovery Service," 64.
 23. Ibid., 81.
 24. Nadine P. Ellero, "An Unexpected Discovery: One Library's Experience with Web-Scale Discovery Service (WSDS) Evaluation and Assessment," *Journal of Library Administration* 53, no. 5-6 (2014): 323-43, <http://dx.doi.org/10.1080/01930826.2013.876824>.
 25. Vaughan, "Evaluating and Selecting a Library Web-Scale Discovery Service," 66.
 26. Hoseth, "Criteria To Consider When Evaluating Web-Based Discovery Tools."
 27. Yang and Wagner, "Evaluating and Comparing Discovery Tools"; Chickering and Yang, "Evaluation and Comparison of Discovery Tools"; Bietila and Olson, "Designing an Evaluation Process for Resource Discovery Tools."
 28. Vaughan, "Investigations into Library Web-Scale Discovery Services"; Vaughan, "Evaluating and Selecting a Library Web-Scale Discovery Service"; Freivalds and Lush, "Thinking Inside the Grid" I Brubaker, Leach-Murray, and Parker, "Shapes in the Cloud."
 29. Chapman et al., "Developing a User-Centered Article Discovery Environment."
 30. Jakob Nielsen, "First Rule of Usability? Don't Listen to Users," Nielsen Norman Group, last modified August 5, 2001, accessed, August 5 2014, <http://www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users>.
 31. Freivalds and Lush, "Thinking Inside the Grid."
 32. Ibid.
 33. Matthew B. Hoy, "An Introduction to Web Scale Discovery Systems," *Medical Reference Services Quarterly* 31, no. 3 (2012): 323-29, <http://dx.doi.org/10.1080/02763869.2012.698186>; Vaughan, "Web Scale Discovery Services"; Vaughan, "Investigations into Library Web-Scale Discovery Services"; Hoepfner, "The Ins and Outs of Evaluating Web-Scale Discovery Services"; Chickering and Yang, "Evaluation and Comparison of Discovery Tools."
 34. Marshall Breeding, "Major Discovery Products," Library Technology Guides, accessed August 5, 2014, <http://librarytechnology.org/discovery>.
 35. Hoepfner, "The Ins and Outs of Evaluating Web-Scale Discovery Services," 40.
 36. Mandernach and Condit Fagan, "Creating Organizational Buy-In," 429.
 37. Bietila and Olson, "Designing an Evaluation Process for Resource Discovery Tools."
 38. Special thanks to Rutgers' associate university librarian for digital library systems, Grace Agnew, for designing this testing method.



Copyright of Information Technology & Libraries is the property of American Library Association and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.