Web Scale Discovery

What and Why?

Abstract

Web scale discovery services for the library environment have the capacity to more easily connect researchers with the library's vast information repository. This includes locally held and hosted content, such as physical holdings, digital collections, and local institutional repositories. Perhaps more significantly, web scale discovery also accesses a huge array of remotely hosted content, often purchased or licensed by the library, such as publisher and aggregator content for tens of thousands of full-text journals, additional content from abstracting and indexing resources, and content from open access repositories.

This chapter defines web scale discovery and highlights a few key concepts essential for understanding these services. For anyone who has worked a reference interview and heard a student utter, "I couldn't find an article in the library catalog," web scale discovery services hold tremendous potential. Extensive research on user expectations in the discovery arena, and the tools used by those seeking information—tools often disassociated from the library and often overlooking much of what the library holds and licenses—provide ample rationale for why web scale discovery is important for the library environment.

What Is Web Scale Discovery?

Connecting users with the information they seek is one of the central pillars of our profession. Succinctly put, Web scale discovery can be considered as deep discovery within a vast ocean of content. The mechanics behind Web scale discovery are not necessarily new, though a commercial application of this approach within the library environment—efficiently and, it's hoped, effectively—is very new. While there are various approaches to Web scale discovery, this issue of *Library Technology Reports* will focus on what today appears to be the most common approach, which, at its heart, involves huge, centralized, preaggregated indexes searched by the end user.

Expanded further, Web scale discovery is—or certainly holds the potential to be—the evolution that libraries have long sought for information discovery. As information professionals, we all have at least a general awareness of the evolution of discovery tools within the library context. Such tools initially were print-based, such as bound handwritten catalogs, the card catalog, and works such as *Poole's Index to Periodical Literature* and the *Reader's Guide to Periodical Literature*. For the past several generations, such tools gradually transferred into the automated, electronic realm, with an obvious example the development and evolution of the online integrated library system (ILS) with a front-end catalog accessible to librarians and end users. These catalogs were initially available within the local library's physical building, often through a menu-driven, text-based interface. With the development in the 1990s of the Web's physical infrastructure and the empowerment of many to access this new environment, library vendors created HTTP Web-based online catalogs. Other evolutions included early pioneers in broad online information systems, including Dialog and LexisNexis. The 1990s ushered in growth in publisher-based electronic journal content, e-text and e-book content, abstracting and indexing databases, and full-text content aggregators looking to pull related information together within an easily accessible and searchable electronic medium. Many of these services or products were initially provided on CD-ROM or through text-based, menu-driven networked systems, which all eventually evolved into

information search and delivery through a Web-based environment.

During this time—we'll say the mid-1990s—many of us witnessed (as we still witness) the confusion suffered by many end users faced with the choice of myriad information systems with their myriad interfaces and specialized bodies of content. Many of us may remember, from around 1998, a new search engine developed by two Stanford graduate students. Who knows whether there was any suspicion that within a decade, Google would often be the first stop—and at times the sole gateway to information discovery—for a new generation of students? Back in the library and academic environment, other information discovery and delivery systems (often separate from the ILS) were developed, such as institutional repository, course management, electronic reserves, and digital collection management systems. These systems provided additional avenues for hosting, discovering, and delivering to end users local library or institutional content. Library vendors developed and marketed federated search solutions, which attempted, with mixed (limited) success, to simultaneously search, retrieve, and adequately display content from various remote information hosts—such as abstracting and indexing and full-text databases—and were often difficult or time-consuming to configure and maintain. In a sense, the original federated search technologies, based on protocols such as Z39.50, represent an early approach to Web scale discovery of content at the article-citation and full-text levels. In the mid- to late 2000s, library Web catalogs evolved into "next-generation" library catalogs, offering increased intuitive functionality and a user interface more in line with popular sites on the Web (e.g., Amazon.com). These next-generation catalogs also provided the capacity to harvest records from various locally hosted library silos of information—such as catalog records from one host and digital collection records from another host—with search, retrieval, and presentation of the collective results from within the single, next-generation catalog interface. In many cases, these next-generation catalogs are still state of the art for many libraries. In short, these systems offered new discovery layer options, uncoupled from any specific underlying ILS. Things began to open up.

This abbreviated and selective history brings us to the present. Library vendors have developed next-generation catalogs with many features and cues understood and expected by today's researchers, such as faceted navigation, tag clouds, and Web 2.0 social features. These systems can resolve a user's research need, in many instances, to the final item level, such as a book. Or they can resolve a user's research need to a collection level, such as the existence of a *Journal of Psychology.* The user is given a call number for a physical copy to browse (if the library still subscribes to print) and a MARC record 856 link, which takes the user perhaps first to a link resolver to select a content provider and then to the publisher's or aggregator's website, where the user must then search again, this time for article-level results, within a different discovery interface. This level of resolution for the library catalog is increasingly not acceptable, as more and more articles from scholarly journals, magazines, and newspapers are available in electronic form. It is troubling that students in 2010 utter the same thing as did students in 2000—"I need an article on psychology, and I can't find it in the catalog." Many libraries have additional avenues beyond ILS-based Web catalogs for further discovery of information in electronic form, such as A–Z lists of the hundreds of databases they subscribe to, A–Z lists of the thousands or tens of thousands of full-text journals large libraries have access to, and librarian-created guides focused on subject areas. As link resolvers and the OpenURL became mainstream in the 2000s, their magic allowed users to resolve their search to full-text articles, though often the user still needed to search a variety of front-end interfaces prior to connecting to the full text article as brokered through the link resolver.

Web scale discovery services for the library environment are an evolution holding great potential to easily connect researchers with the library's vast information repository. By preharvesting and centrally indexing content sourced across multiple silos, Web scale discovery services hold the promise to fundamentally improve and streamline end user discovery and delivery of content. Such content includes physical holdings, such as books and DVDs; local electronic content, such as digital image collections and institutional repository materials; and remotely hosted content purchased or licensed by the library, such as e-books and publisher or aggregator content for thousands of full-text and abstracting and indexing resources. For purposes of this issue of *Library Technology Reports,* Web scale discovery can be considered a service capable of searching across a vast range of preharvested and indexed content quickly and seamlessly. Web scale discovery services provide discovery and delivery services that often have the following traits:

- **Content.** These services harvest content from local and remotely hosted repositories and create a vastly comprehensive centralized index—to the article level—based on a normalized schema across content types, well suited for rapid search and retrieval of results ranked by relevancy. Content is enabled through the harvesting of local library resources, combined with brokered agreements with publishers and aggregators allowing access to their metadata and/or full-text content for indexing purposes.
- **Discovery.** These services have a single search

box providing a Google-like search experience (as well as advanced searching capabilities).

- **Delivery.** These services provide quick results ranked by relevancy in a modern interface offering functionality and design cues intuitive to and expected by today's users (such as faceted navigation to drill down to more specific results).
- **Flexibility.** These services are agnostic to underlying systems, whether hosted by the library or hosted remotely by content providers. These services are open compared to traditional library systems and allow a library greater latitude to customize the services and make the service its own.

Why Web Scale Discovery?

As illustrated by research dating back to as recently as 2010 or as far back as the 1990s (if not earlier), library discovery systems within the networked online environment have evolved, yet continue to struggle to serve users. As a result, the library (or systems supported and maintained by the library) is often not the first stop for research—or worse, not a stop at all. Users have defected, and research continues to illustrate this fact. Rather than weave these research findings into a paragraph or page, some illustrative quotes convey this struggle. Those wishing to read the full context of the research can do so at their leisure. The quotations below were chosen because they succinctly capture findings from research involving dozens, hundreds, and in some cases thousands of participants or respondents:

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

> * * *

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

* * *

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

> End users' expectations of data quality arise largely from their experiences of how information is organized on popular Web sites. . . . [User] expectations are increasingly driven by their experiences with search engines like Google and online bookstores like Amazon. When end users conduct a search in a library catalog, they expect their searches to find materials on exactly what they are looking for; they want relevant results.[4]

* * *

Users don't understand the difference in scope between the catalog and A&I services (or the catalog, databases, digitized collections, and free scholarly content).[5]

* * *

It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of "silo" systems whose distinctions seem arbitrary. . . .

> The continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings. Our users expect simplicity and immediate reward and Amazon, Google, and iTunes are the standards against which we are judged. Our current systems pale beside them.[6]

* * *

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

A: Just remember that students are less informed about the resources of the library than ever before because they are competing heavily with the Internet.[7]

Other factors, apart from user behavior and preferences, also give reasons for libraries to use Web scale discovery services. First and most obvious is that if something is not discovered, it has no chance of being used. Whether a librarian conducts a reference interview, a user browses the shelves, a friend provides word of mouth, a user searches in Google or a library database, or a user scans issues and article titles in an electronic journal, discovery must happen, either by focused intent or serendipitously. Libraries

often spend tremendous amounts of money every year to purchase or pay for access to an ever-growing body of electronic content, and the cost for access to this content often increases on an annualized basis. At least for academic libraries, most collections monies often go for electronic, not print, resources. Clearly purchases are made with the hope that the content will be used. For the content to be used, it must be discoverable—and for many of today's users, easily discoverable. Publishers and aggregators also have a vested interest in discoverability and use of their content. Libraries regularly review content-usage statistics and frequently analyze overlaps within their collections. Extensive research into whether these Web scale discovery services increase discovery and usage of publisher and aggregator content does not yet exist, though it seems reasonable to assume that if materials are more easily discoverable, they will be used more heavily and access statistics will increase. Statistics are not the sole factor libraries analyze in deciding whether to continue a subscription, but they often play an awfully big role. While extensive research doesn't exist, some studies have been published. For example, Doug Way compared database and full-text usage statistics before and after the implementation of the Summon service at Grand Valley State University Libraries.[8] His research, involving analysis of GVSU's link resolver statistics before and after Summon's implementation, suggests that the new discovery service was broadly adopted by his institution's population and has resulted in an increase in the use of the library's electronic resources. Bill Kelm at Willamette University has presented results indicating that Willamette's implementation of WorldCat Local has similarly led to an increase in the use of the library's electronic resources and the number of ILL requests.[9] Ex Libris Primo Central, Ebsco Discovery Service, and Innovative Interfaces Encore Synergy were all released in 2010, and no published research has yet been identified, at least in reference to surfacing Web scale content as offered by these services.

Another tie-in focuses on information literacy. To the degree that these new tools may be used as a gateway to information by users, such use could serve as another tool to build information literacy. Web scale discovery services are by default searching local library collections and remote licensed content, purchased (it's hoped) with a purpose. Presumably, this information, at least collectively, is more accurate, relevant, and appropriate for research purposes than, for example, a webpage with dubious, less informed, or opinionated content that may appear on the first page of a search engine's results. In other words, the content is vetted to some degree as more likely to be worthy and true. At least one discovery service has a focus on scholarly articles, and several services allow users to refine a search to only materials designated (by Ulrich's or through a separate analysis) as scholarly or peer-reviewed. One service has an optional scholarly article recommender component, which helps bring together related scholarly materials in an increasingly interdisciplinary research realm. In short, if part of building information literacy involves steering users to discovery systems exposing library-vetted content, and if the features and content scope of such tools are attractive enough for users to bite, then these new discovery tools hold great potential.

Audience and Scope

The primary goal of this work is to provide a valuable foundation to libraries that wish to know more about library-focused Web scale discovery services and to aid libraries contemplating a marketplace review for their local environment. It does not presume that the reader is familiar with any of these services, nor will the report delve into extreme technical detail. Such detail is best addressed by direct library-to-vendor dialog because that approach will supply the most up-to-date information available and because libraries may often have specific questions pertaining to their unique position or environment, local library skill sets, and staff workflows. If this issue of *Library Technology Reports* introduces a new concept to some readers, serves as a foundation for other libraries to utilize in their own marketplace evaluations, or provides a few additional questions or twists that those already well along in their investigations hadn't considered, then it has met its goal.

The author identified five Web scale discovery services from major vendors: OCLC WorldCat Local, Serials Solutions Summon, Ebsco Discovery Services, Innovative Interfaces Encore Synergy, and Ex Libris Primo Central. This report profiles four of the five. Encore Synergy by Innovative Interfaces is not profiled, which in no way implies that this product is not of interest or should not be considered by potential library customers. Indeed, Encore Synergy is already in use at a host of academic and public libraries. However, given space restrictions, time restrictions, and the fact that Synergy was one of the most recently released services, boundaries had to be drawn somewhere. In addition, the approach of Encore Synergy to Web scale discovery is a bit different from that of each of the other four services in this report; Synergy does not create a preharvested, preaggregated central index for local and remote publisher or aggregator content. Like other vendors, Innovative Interfaces has made agreements with content providers, though its approach focuses on using modern Web services to access the publisher or aggregator content, remotely hosted, in real time. This author makes no judgment calls on whether this approach is better or worse than others, and potential

library customers owe it to themselves to investigate each of these five services.

Another boundary is that this report focuses on released, publicly available products sold by library vendors to the library community. It does not focus on the development of open source initiatives related to Web scale discovery of library resources, such as the quite interesting eXtensible Catalog (XC) project spearheaded by the University of Rochester.

A Few Key Concepts
The Discovery Layer and Harvesting

Continuing the trend that next-generation library catalogs introduced to a wide audience a few years back, Web scale discovery services unlatch the discovery layer—the interface within which the user works for information discovery and delivery—from the traditional underlying host and delivery systems maintained at the local library level. Discovery services harvest or otherwise obtain ILS catalog content and content from other local repositories, such as digital collections and institutional repositories, often through the OAI-PMH protocol, FTP, or a related mechanism. For ILS records, the level of content is adequate so that a user looking at the record in the discovery system generally need not click through to an underlying full catalog record as housed in the ILS, especially considering that enriched content and real-time status calls are integrated into the Web scale discovery experience. Still, each service provides a link to the full catalog record within the native ILS interface. For other materials, the discovery service may surface the item and provide a link to a more capable, function-replete system housing the native content. For example, digital collection management systems often include additional functionality that at present isn't offered through discovery service interfaces, such as zooming, panning, and rotation of images. It's important to note that the descriptive staff work related to the library's print and e-content collections—the original cataloging and other metadata work—is still performed within the underlying library system, whether it be the ILS, digital collection management system, or institutional repository or related system. Staff are not natively cataloging records within the discovery system database/interface. In addition, local library systems may provide additional indexes and associated fielded searches not offered through the Web scale discovery service directly. In summary, at this stage of development, there is a role, from the perspectives of the user and library staff, for both the encompassing Web scale discovery service and the user interfaces and underlying systems associated with the library's local collections. Similarly, publisher and aggregator content may often reside in a database tuned to that particular content type or for a particular audience or discipline that typically uses that resource (e.g., Medline). Therefore, there continues to be a need for the end user to be able to utilize the native interfaces of these resources. Recognizing this need, several discovery services have recently introduced or begun investigations into database and collection highlighting within Web scale discovery results. For example, for a broad search, "engineering," the database IEEE explorer may be returned at the top of the list; the user can click this link to connect to and search the IEEE database directly, leaving the web scale discovery service.

As mentioned, each discovery vendor profiled in this report has entered into agreements with content providers that allow it to harvest, normalize, preindex, and expose through its Web scale discovery service the publisher or aggregator content. To facilitate regular, automatic content updates as new information is published, automated transfer routines and load tables are configured between the discovery services and the content vendors. Similarly, routines and procedures are set up between the local library and the discovery service vendor to accommodate updates and changes for the locally harvested records (additions, deletions, changes). Harvested content is normalized into an underlying schema, developed by the discovery service vendor, that facilitates indexing, relevancy ranking, and an even level of presentation for different content types with potentially varying levels of metadata.

Exposure of Licensed Content

A key concept is the exposure—or surfacing—of licensed content within the discovery service indexes. As a test-drive of some of the already live and searchable library-specific discovery services—as well as Google Scholar—will easily illustrate, a library need not be a subscriber to much of the content contained in the discovery service index for the content to be discoverable. In other words, in many instances, the user is actually searching a broader collection than the local library may physically possess or own or license electronically. This is a potential boon to the researcher. Simply knowing of an item's existence is better than not knowing, assuming that appropriate delivery options (e.g., interlibrary loan) are available if the library hasn't licensed access to the electronic full text. In general, access to the full text is where a user must be authenticated. That said, uniquely indexed content, in particular content from some abstract and indexing databases, is not open for surfacing unless the library also subscribes to that resource. However, citation-level content covered in subscription abstracting and indexing databases is often available directly from publishers, aggregators, or other databases. Put another way, even if one didn't consider *any* of the

e-journal, publisher package, or database subscriptions and licenses the library pays for at its local institution, there is a very large base of content indexed by the discovery services that is available for search and discovery. If a library had no paid e-content licenses or subscriptions, there would still be a large index of content to search against. Publisher agreements can permit these discovery services to index their content and provide access to citation-level metadata within the central index to all customers regardless of whether the local library itself has a licensing agreement with that publisher and has purchased access to this content. When a user clicks through to access the full text (usually using a link resolver and associated rights management information), that's usually the point where authentication will be requested (assuming a user isn't already authenticated; if a user is physically at a library site or on campus, authentication is usually behind the scenes and automatic, based on authorized IP address ranges tied to the library subscription). In addition, each of these discovery services includes a huge amount of open-access content. If a picture is worth a thousand words, then a test-drive of the very open Summon index will quickly illustrate this concept. You, as an unauthenticated user with no known association to a library or a library's e-content subscriptions, can, from off site, go to any Summon library website, leave the search box blank, and execute a search. Then click on the refinement option Add Results Beyond Your Library's Collection. Enjoy the half billion plus items that have just been surfaced. Of course, it would be frustrating if there weren't a method available to scope to the local library collection—the physical collections and licensed e-content accessible to that library's population. Fortunately, each discovery service is attuned to this issue and has one or more ways of providing just results for which the user has rights to, and/or providing a visual cue—whether it be a search that is by default scoped to the subscribing library's collection (which users can expand if they wish), a Full Text Available icon next to the brief item result, or some other approach. In some cases, full resolution to an item for a library may be an interlibrary loan request, and discovery services provide for this capability as well, often in conjunction with the library's link resolver knowledge base.

A Note of Caution

Web scale discovery services, as defined in this report and focused on the library environment, are extremely new. After development periods for each service, OCLC WorldCat Local became publicly available at the end of 2007, Serials Solutions Summon in January 2009, Ebsco Discovery Service in July 2010, and Innovative Interfaces Encore Synergy and Ex Libris Primo Central both in mid-2010. In short, the majority of services, at press time, are a year old at most. Each service is evolving extremely rapidly, with enhancement cycles measured in months, if not weeks. During the writing of this report, things changed rapidly, and significant revisions occurred between drafts. As the report goes to press, vendors will continue to ink new contracts with additional content providers, adding more content to an already huge pool. There will be interface updates, changes, new capabilities, and cosmetic changes. That said, the purpose of this report is not to highlight, let alone mention, every last detail, capability, and set of information associated with each service—that is beyond the scope of a single edition. This report profiles four discovery services; in the past, a whole issue of *Library Technology Reports* has focused solely on a single service.[10] The purpose of this report is to provide a broad, yet substantive overview of several of the major services. Through presentations and conversations, the author has found that awareness of these tools in the library environment isn't necessarily widespread yet. In short, this issue of *Library Technology Reports* is meant to introduce or further introduce the concept of Web scale discovery, to profile and provide an overview of several of the key services available today, and to provide some questions for thought and additional resources that potential library customers may wish to consider in the context of their own evaluations.

Given the quite extensive array of features and functionality associated with these services, the extreme pace of development, and the flexible nature of these services, which allows a high level of local library customization and integration within the library discovery environment, there may be an unintended inaccuracy or two, or otherwise outdated information somewhere within this work. The author has made his best efforts to report accurately and apologizes in advance if errors are discovered. Research efforts related to this work included identification and review of existing publicly available and published information on these services—vendor website information and press releases, webcasts sponsored by vendors or third parties, and other published literature (which remains somewhat sparse at this early stage). It also included test-driving multiple beta or live sites for each of the profiled services, with an eye toward what the end users see. It also included detailed questions sent to each vendor, and often additional follow-up. If a particular feature is mentioned or highlighted for one service and not others, readers shouldn't automatically assume that the other services do not have the same or a similar feature or capability. It may mean that this feature jumped out at the author while he was preparing one of the particular profiles. Or it may reflect the fact that vendors often answered the same question with different information, at different levels

of completeness, or from a different angle and focus. This should be no surprise to those who have conducted comparison research in the past (this author included), looking at different tools performing similar functions. For reasons that should be obvious, it would be foolish to take each new angle or feature mentioned by one vendor for its service and ask all the other vendors about it. If this approach were pursued, this report would turn into a truly Sisyphean task and would never reach completion—by the time one draft was done, each service would have changed, offering new features, and the cycle would start anew.

It's fair to say that during the course of this research, vendors at times challenged publicly made statements from other vendors[11] that, for example, their service is the only one that has this feature or is the only one that covers this content. The author has striven to write this report in a neutral tone and hopes no bias toward one product or another is present. This report specifically does not include a comparison matrix of features, functions, and content for the services. Given that things are changing so rapidly with these tools, and given the reality that quite a lot of local configuration options are available to each library customer, a comparison matrix did not seem appropriate. Within each of the four chapters profiling individual discovery services, a short Vendor Perspective section is included. The author invited each discovery service vendor to provide its own short sound bite highlighting its particular offering. Each vendor chose to participate. The author suggested to each vendor that it briefly touch on the following broad points related to its Web scale discovery service:

- **Content scope—**the capability to present local library content, such as ILS content and local digital collections, as well as, and with a special focus on, the centralized preaggregated index of primary and secondary publisher and aggregator content
- **End-user searching and functionality—**highlights about the user interface
- **Local library customization capabilities—**highlights related to how the library can make the discovery service its own through customization, choosing what features are offered, branding, and so on

Ground rules were simple and direct. Vendors were asked to focus on their service and why they think it is a worthy contender and not focus on competitive marketplace offerings. They were asked to keep their responses neutral in tone and were not allowed to mention competing vendors by name. In short, they were given an opportunity to succinctly summarize their service and what they think makes it stand out. The target word count was 500, and each vendor came in at around that

length. The author has chosen not to edit the Vendor Perspective content, and thus the "Vendor Perspective Section" in each chapter is truly the vendors speaking in their own words.

From one perspective, perhaps the most delicate topic within this report is that of publisher and aggregator indexed content. All vendors have made and continue to make agreements with publishers and aggregators for the rights to index content for purposes of surfacing such content within the discovery service. The exact scope of the content associated with each service is challenging to nail down, and vendors have provided varying levels of information on this topic; for some content-related questions, they chose not to provide particularly illuminating responses—somewhat understandable, given the competitive marketplace and, perhaps in some cases, the confidential nature of signed content agreements. For any specific publisher or aggregator, a different level of content—basic metadata, more detailed metadata, author-provided metadata, abstracts, full text—may be provided to the different discovery vendors. Some vendors may have additional in-house staff who enrich content that's been received. Some discovery vendors may have exclusive agreements with one or more content providers. Content for the same item—a single unique journal article, for example—can be sourced from many different outlets. The amount or richness obtained from one outlet may not be equal to that from another outlet. In short, the devil is truly in the details.

Each library considering a serious marketplace review should thoroughly do its homework, which in this case means thorough questioning of the vendors. It also means conducting plenty of sample searches in existing live implementations, perhaps focused on the subjects or material types most pertinent to your clientele and analyzing the results. It also means conducting your own assessment of existing usability studies and published literature. To borrow a direct quote, as the author doesn't believe he can more succinctly express the same sentiment, with which he wholeheartedly agrees, "As history has shown, multiple solutions arise to address real needs, and each solution has its own characteristics. In terms of discovery solutions, I'm confident that each library, after conducting a thorough evaluation of facts and features, will be able to determine which of the available products best fits the library's mission, needs, policies, and environment."[12] The author hopes this issue of *Library Technology Reports* serves as a bright torch to help light your way as you consider, begin, or continue investigations into Web scale discovery services within your own library environment.