11-2013

Web-Scale Discovery and Federated Search

Valeri Craigle

Law Libraries in the Digital Age Chapter 8: Web-Scale Discovery and Federated Search

By Valeri Craigle

1. Introduction/Overview/Purpose

In stark contrast to the library card catalogs of old, today's library search interfaces offer much more than one-dimensional, item-specific searching. Users are now engaged in a process of discovery in which they are empowered to control not only the sources of content being searched, but also the context into which information is delivered, and the platform onto which information is synthesized. By eliminating the barriers to information discovery, law libraries can position themselves as true partners in this process, defining their mission in new ways, and providing critical services in an ever-complex information ecosystem.

Discovery technologies are vital to a law library's core services. Users are driving the demand through a now commonly held expectation—largely influenced by Google—that searches should include "everything" the library has to offer and should be displayed in a centralized location with a variety of options for organizing and contextualizing the information. Law library users are no exception, though by the very nature of the current legal information environment, legal researchers must have a certain tolerance for information silos such as Lexis and Westlaw.

As Discovery technologies take hold in the legal information environment, law libraries need to understand the inner workings of these systems and the roles that must be fulfilled in order to meet the needs of their users. As discovery search services continue to make rapid progress, understanding how they work, and having an awareness of the implications of deploying these services in law libraries, will keep law librarians ahead of the curve.

2. From Search to Discovery

Legal Research Beginnings

Legal research is unique from the research process of any other academic or professional discipline. Practitioners must navigate a complex tangle of information scattered among a multitude of locations, formats, and iterations. Skill and experience are required to master the landscape. Core legal research in case law, statutory law, and other primary materials is generally conducted in the native environments of large and complex databases like Westlaw, Lexis, and Bloomberg Law. The use of print resources has dropped dramatically in the age of electronic databases, or what has historically been referred to in the legal research pedagogy as Computer Assisted Legal Research (CALR).

CALR resulted in a sea change for the legal profession. Development of the earliest legal information databases began in the 1960s and '70s out of a very simple need to more quickly and comprehensively search the exploding body of legal knowledge. There was just simply more law to research and it was growing at an exponential rate.

One of the first legal research databases developed was the "Horty Project". Known by some as the precursor to Lexis, it was an initiative started in 1959 by John F. Horty, a professor at the University of Pittsburgh. The purpose of the project was to compile Pennsylvania state statutes to facilitate legal research in health care law. In 1971, the "OBAR" Project (Ohio Bar Automated Research) launched to deliver the entire collection of Ohio state codes and court decisions. Soon after, in 1975, Westlaw came onto the scene with its famous reporting system[1].

By the early '90s, Lexis and Westlaw were evolving into massive databases, developing large-scale business models and increasingly proprietary philosophies. The pro se movement

spurred the evolution of online free legal research resources, even when the Internet was relatively new. The Legal Information Institute, started in 1992, offered a host of primary legal materials, both state and federal, also offering a wealth of secondary sources, particularly in the areas of legal ethics and Social Security.

Today, there are too many free online legal resources to count. Google offers a search engine designed specifically for case law. FindLaw offers cases and codes, Supreme Court decisions, and lawyer directories. And many websites have begun offering free legal advice, or advice offered for a minimal fee. While legal practitioners and law schools continue to subscribe to the large, fee-based resources like Lexis, Westlaw, Heinonline, and Bloomberg Law, the growing pro se phenomenon is driving the development of free online resources and self-representation collections in law libraries. As the economy continues to stagnate and the cost of legal resources remains prohibitive for much of the population, the growth of free, web-based resources and discovery tools will only increase over time.

But free legal resources found on the web do not offer the one thing that large database publishers do: the ability to organize legal information into a meaningful context. Legal research is not just a process of discovery, but of contextualization among related resources, and resources which provide deep analysis of a particular point of law. Serious researchers are looking for "good" law, not just anything they find on a particular subject matter. The big legal databases offer this perspective and contextualization, but they charge a high price for it, and they are only available to a select few.

The Current Environment

With so many high-tech tools at our disposal, one would expect the legal research environment to be nimble, seamless—practically automated. In actuality, the tools we use are more numerous and complex than ever. Most primary legal sources are siloed away in large, proprietary databases, accessible only to those of us willing to pay top dollar for high-end search algorithms and an endless array of tools and features.

The corpus of the law has grown in such size and scope that we need road maps to navigate the convoluted road of legal information. And as the size and complexity of primary legal resource discovery grows, so does the discovery path to secondary sources such as law reviews, journal articles, commentaries, and news stories, which serve as vital sources of analysis, and legal definition and description. For today's legal practitioner, knowing how to find secondary source material is a critical skill, particularly for new lawyers, who spend a large amount of time in their first few years of practice doing secondary source legal research[2].

Discovery platforms and federated search are emerging as effective search solutions for research into secondary source materials, particularly web-based resources. As these systems gain recognition in the legal research community, law libraries must understand what products are available, how they function, and how to be proactive in the implementation and management of these systems.

3. Drawing distinctions

Federated Search

Federated search is a search technology that provides a one-box search experience for the user, similar to google, that searches multiple library databases at the same time. In this way, the patron doesn't have to know whether they need to search Heinonline or LLMC or Proquest Congressional to get their information needs met, they can just search once and find results.

The technology returns a results set and users can toggle between sources that were searched. The down side has to do with the fact that relevancy ranking is not really possible because each database uses its own algorithm for determining relevancy.

Similar to discovery platforms, federated search results are displayed in a unified interface in a web browser. But unlike web-scale search, federated search queries databases one at a time, instead of drawing from a pre-existing index like Web-Scale and Discovery Platforms do. Not surprisingly, , federated searches are slower and less efficient compared to that of web-scale search services. Federated search systems are known to frequently time out, particularly during times of high use or if multiple authentications are required. As a result, libraries have largely abandoned federated search in favor of web-scale discovery, a suprising development considering how popular federated search systems were just a few years ago.But federated search is still utilized in certain sectors, primarily by science and technology researchers for searching the deep web[3]. And although federated search is no longer preferred as a discovery solution for libraries, it still offers a comprehensive, real-time search across multiple resources, without relying on pre-existing indexes, which may not be updated on a daily basis.

Discovery Layers

A discovery layer is simply the public interface to library resources cooperatively selected by a library and discovery layer vendor for discovery in the system[4]. These resources may include anything purchased, licensed, indexed in the catalog, or living out on the web. You could think of a discovery layer as a next generation OPAC, only much more robust, with lots of

search and browsing features, and some social media tools. It's likely that if you have done any searching in an academic library catalog in the last few years, you have used a discovery layer.

In discovery layers, researchers don't just search for information, they interact with it and can modify it to suit their needs, creating a personalized research experience. A medical researcher can limit search results to peer review articles only; an art history major can search for only audio-visual items, then limit the location of those materials to a digital collection; a law librarian can write a review for a resource he has found to be valuable first year legal research students.

One of the greatest advantages of a discovery layer is that it provides a unified view of all the items contained in the vast repository of library resources. Bibliographic and other descriptive information is gathered from all of these sources and placed in one location for searching, called the Central Index..Unlike federated search which sends out queries to all the different databases and then waits for all the results to come back, the discovery platforms have pre-indexed all of those sources instead. That way everything is in the same central database or index and can be quickly searchedAlthough there are a number of discovery layer platforms on the market, law libraries seem to prefer Encore as the platform of choice, according to an informal survey of technical services law librarians in 2010.52 percent of the librarians surveyed indicated that their libraries were either in process of implementing discovery layers or had already done so. Encore was by far the leading product, with WordCat Local and Aquabrowser coming in second and third, respectively[5].

Other discovery layer products include:

- Primo (Ex Libris)

- Enterprise (SirsiDynix)
- Ebsco Discovery Service (Ebsco)
- Summon (ProQuest)
- VuFind (Villanova University)
- Endeca (Oracle)

Web-Scale Services

In the last few years, there has been a lot of focus on more effectively managing user traffic to databases, particularly during times of heavy usage, which tends to slow down a web service. Before web-scale discovery services, heavy traffic would cause these systems to crash altogether. When a system is "web-scale," it means that it is built on an architecture that balances and stabilizes a system when a lot of people are trying to access it at one time. When a web-scale system detects an increase in traffic, it slows down the parts of the system that are not being used so that more resources can be allocated to the parts of the system that are being heavily used. balances, [6]. Web-scale services are a relatively new technological innovation, whose development gained traction in the early 2000s[7]. Web-scale services slow traffic just enough to keep the data flowing, then return to normal functioning when the loads are not so heavy. This results in consistent performance over time, ensuring a constant flow of data and few system failures. In many libraries today, web-scale services form the backbone of search services, online bibliographic utilities, data management systems, and patron services. Indeed, web-scale services are at the functional core of discovery layer platforms and online search engines.

4. Inner Workings of the Discovery System Platform

Technical Infrastructure:

Discovery layer platforms contain several components, all working in concert to bring this new search experience to the user. Hardware, software, hosted services, and web-scale services each play a role in the proper functioning of the system as a whole. Each discovery layer vendor may have a unique approach in the deployment of these services. Some may require all components to be purchased and held locally, some may choose a hybrid solution of local hardware and some web services. Others may deploy their services exclusively from the cloud.

In terms of hardware, vendors may ask a library to purchase a number of data servers to perform such functions as data backup, data testing and staging, and data harvesting.. Often with discovery layer platforms, the central index is stored in the cloud to take advantage of web scalability, so the servers that the library hosts are needed mostly to boost performance and provide extra data security. Some vendors will give libraries the option to have these data services hosted by them if they lack the budgets or personnel to administer the servers themselves.

Software components include the discovery layer interface itself; web-based administrative interfaces for controlling the look, feel and functionality of data; mobile applications; and statistical software for analyzing user behavior, collection use, and any other reporting needs a library might have. Law libraries may find great value in the reporting software when submitting their annual ABA or *U.S. News* collections statistics. Again, these services can all be deployed in the cloud or from a desktop client.

Ideally, all components would be deployed in the cloud to provide the scalability, flexibility, reliability, and seamless delivery required by the large data loads and numerous

services necessary for discovery layer platforms. As these systems evolve, the need for them to be web-scalable will only increase.

Information Gathering, Management, and Display

As previously mentioned, one of the greatest advantages of web-scale architecture is its ability to quickly return search results by drawing from a pre-harvested central index. This index contains resource metadata as well as authentication information for access to full-text materials. The metadata for databases, e-journals, and other publications is supplied by publishers who have entered into cooperative agreements with the library implementing the discovery layer platform. Metadata for digital collections, Marc metadata from the catalog, and metadata from external data sources must all be gathered along with publisher's metadata in the central index.

Once metadata from these various sources is placed in the central index, a series of customizations can be performed for the visual display of items in the discovery layer interface. Customizations can be made to features like browsing facets, social media tools, and even book icon images. . Customizations can also be made for simplifying the display of items such as books that have multiple copies, locations, or formats. Permissions for performing these customizations is usually limited to server administrators. As patrons begin to interact with the system, adjustments may need to be made periodically to accommodate user preferences. For example, people may want more facets in a particular subject area, or requests might be made to display book chapter icons in the results list. For academic law libraries, campus-wide discovery layer implementations, which unify the resources of the entire campus, may make legal research more complex, adding to the mix resources from hundreds of disciplines. Complicating things further is each party's definition of "relevancy." What is relevant in the biological sciences may

not be relevant in the humanities, the arts, or the law. Who decides which results float to the top as more or less "relevant"? These issues must be resolved cooperatively among the campus libraries participating in the implementation, with the caveat that there may not be a perfect solution for all of the parties involved.

Search Results and Access

The first encounter with a discovery layer interface can be daunting. Users are often confronted with an overwhelming set of choices. Books, articles, audio files, digital photographs—almost any material type one can think of appears in the search results screen. Narrowing a search requires some skill, challenging the user to utilize a series of facets, filters, sorting options, and other advanced search features. Law library patrons must learn to master these tools in order to limit their results list to a manageable size. Law librarians will discover that suddenly their patrons require a bit of training to navigate the complexity of this new environment. Legal researchers may become frustrated by this new complexity and wish for the simplicity of the old catalog.

Indeed, it raises the question of whether it is beneficial for law libraries to participate in campus-wide implementations, as flooding the legal information landscape with materials that are irrelevant to legal researchers may be counterproductive. However, discovery platforms can be extremely beneficial for cross-disciplinary research and research involving secondary sources in the law. Materials that the researcher may never have thought of might appear. Environmental law researchers could discover water-rights maps from a digital collection, or an article in a scientific publication about the detrimental effects of pesticides. By expanding their horizons, discovery platforms open a new world to legal researchers.

For the patrons unaffiliated with the law library, and thus excluded from access to legal databases, a discovery layer search presents a litany of results for which they haven't the privileges to use. Some publishers prohibit any distribution of full text to non-law patrons, even if they come to the library personally to work with a librarian to gain access to these materials. Law libraries must decide what their policies are for making restricted information visible in the discovery layer. Sometimes it's a matter of choosing the lesser of two evils: whether it's better for patrons to know about a resource and not be able to access it, or whether it's better for them to be unaware of these resources entirely, so as not to cause frustration when they try to acces these materials.

5. The User Experience

Relevancy

Displaying search results according to relevancy ranking is one of the hallmarks of a discovery layer platform. Though the concept is not new, recent advances in the development of relevancy ranking algorithms form the basis of search results display in next generation catalogs. The underpinnings of relevancy algorithms are generally proprietary and are thus poorly understood. Factors such as currency, frequency and proximity of search terms, as well as user preferences and behavior are all part of the formula. Control of these algorithms is generally limited, but some vendors allow libraries to disable relevancy as the default display for search results.

Social Media

Have you ever wanted to post a book review, add a subject heading to a tag cloud, or "like" an article for your friends to see on Facebook? How about tweet a Supreme Court opinion,

or select a novel for your Pinterest page? For the social media native, these might be perfectly natural activities, to be relished as part of the research process. And today's discovery layer platforms offer every social media tool you can imagine to indulge this new appetite for participation.

But one has to wonder if the incorporation of social media tools in the library catalog is simply a marketing ploy, or if it really is a serious attempt to involve researchers more in the discovery process. One could possibly view some of these activities as akin to metadata crowdsourcing. What is a tag cloud, but a cluster of subject headings? Does "liking" an item increase its relevancy rankings? Will book reviews improve descriptive metadata? Will folksonomies replace the subject librarian? Only time will tell if these tools are nothing more than another set of superfluous features. But this is a time when patron-driven acquisitions and metadata crowdsourcing are becoming more accepted as legitimate tools to enhance library services. The Google experience has been a wake-up call for librarians, who can no longer think of themselves as the gatekeepers to information procurement and access. It will be interesting to observe the impact of social media tools on patron interaction in the Discovery Layer platform.

6. Implementation

Cost and Commitment

The cost and commitment of a discovery layer platform implementation is concomitant with the size and complexity of the library and the scope of its collections. Independently operating law libraries who are able to work with their current ILS vendors and electronic resource publishers, and who have no additional digital collections, e-publishing services, or deep web searching needs, can look forward to a relatively smooth process. Any costs incurred

will likely be the vendor's time in loading the central index and working with librarians to develop the appropriate display options for the discovery layer interface.

That is the best-case scenario. Another scenario is an academic law library embarking on a campus-wide, multi-library implementation. If this entails migration of the catalog and the purchase of a new suite of tools that are more compatible with the new discovery layer platform, the process can be costly and time-consuming. In this scenario, libraries should plan on at least 6 months to a year for the catalog migration alone. Additional time and resources may be required to bring into compliance any external systems that will need to work with the discovery layer, such as an automatic retrieval system, or other systems that manage patron requests and circulation processes.

Understanding the full scope of the data to be harvested for the central index is paramount. A large digital collection, for example, may require additional servers or server partitions. Anticipate the need to purchase either additional servers, or bump up existing systems to handle these loads. During initial implementation, a significant amount of time is allocated to systems administration and the participation of technical services librarians. Trainings, meetings with the vendors, and hands-on working sessions to fine-tune customizations will monopolize staff time for the first six months. Plan accordingly for these obligations.

Once the system is operational, downtimes for the catalog will need to be strategically planned. Cataloging lockouts of potentially several weeks may be required for loading the indexes and assessing data compatibility and integrity in the new system. It is during this time that unexpected issues occur. Data could be corrupt, peripheral systems might be incompatible,

or server capacity may be deemed insufficient. All of these factors must be taken into account when planning for a discovery platform implementation of this scale.

Trends

Discovery layer platforms have not been around for very long, only really gaining traction since 2009. A conversation of how discovery layers were being used in law libraries was held during a presentation on discovery layer technologies given at AALL in Portland, Oregon. At that time, very few law libraries seemed to be participating in these initiatives.

An informal survey of technical services law librarians showed that a little over 50 percent of law libraries had implemented or were in the process of implementing discovery layers in their libraries[8]. Though the survey did not address the question of the type of libraries responding, the comments suggested a high number of academic libraries who were participating in campus-wide implementations. Law Librarians try diligently to stay abreast of the latest technologies and want to offer their patrons the best services available. As discovery layer platforms gained traction, Law Librarian's interests were piqued by the prospect of simplifying the legal research process.

In the early days of implementation, most legal information sources were nowhere to be found in the knowledge bases of discovery layer platforms, a common complaint among academic law librarians. Partnerships between legal information publishers and discovery layer companies were slow to form. Today, the availability of these resources is not as much an issue as gaining full-text access to law library resources for patrons who are not affiliated with the law library.

This raises the question of whether it is prudent to enable these resources in a discovery layer at all. If the majority of the patrons who use the catalog can't get access to these materials, does this unduly frustrate the process? Alternatively, are legal researches at large academic institutions frustrated by being flooded with resources that have nothing to do with the law? And does this feedback have an impact on other librarians' decisions to implement this new technology?

For solo law librarians or law firm librarians working in discovery platforms that manage materials specific to their organization, the problems of discipline-specific heterogeneity and resource access don't apply as much as they do for law librarians at large academic institutions. In fact, discovery solutions may be overkill for small libraries, as they often subscribe to much fewer databases and other sources for inclusion in their catalog. By allowing their patrons to "discover" resources that they can't get full-text access to, may aggravate their patrons more than offer a good service If academic law librarians are going to spend a great deal of time, effort, and money to implement a discovery layer platform, they will want a quality product and rewarding experience for their patrons, not a product that leads to more complexity and confusion.

7. The Future

Predictions on The Cloud

In today's nimble technological environment, mobility and platform agnosticism is key. Using local hardware and software to run applications is a paradigm that has run its course. When the concept of cloud-based services started taking hold around 2011, libraries were quick to harness the power of this new technology. In fact, libraries were pioneers in the process, harnessing the power of the OCLC collective in copy cataloging services. Libraries, like many

organizations both in the private and public sectors, have begun to relinquish the acquisition and control of the hardware and software that formed the backbone of data management and dissemination. The cloud offers many benefits for managing the various functions required for information management. As with any system, there are disadvantages to be wary of, but these are far outweighed by the benefits of cloud-based services.

Scalability, resource aggregation, and the unified search interface have transformed Library services. The vast amounts of data libraries manage today and the user loads systems experience while patrons access this data 24/7 will keep library services in the cloud for decades to come. The concept of information silos will become a thing of the past. Data providers, even the most restrictive legal publishers, will bow under pressure from their customers to conform to this new model, exposing their metadata to cloud-based aggregators for harvesting to discovery platform providers, which will become ubiquitous in law libraries in the next decade.

As libraries move away from the idea of local control of resources and turn over the management of their systems to remote providers, they will continue to rely on the cloud to provide these services. All indications point to use of cloud-based computing well into the next decade and beyond.

Next generation Frameworks

The culture we live in today, though tech savvy and highly sophisticated is fraught with information overload[9]. It is no wonder that finding the best formula for information discovery has become one of the hottest topics of our time. Information and data are now valuable commodities, to be bought, sold, and traded. The value of data is compared to that of oil or gold. Personal data, the sine qua non of the information economy, is exchanged by governments,

companies, and marketing bodies, unbeknownst to most Internet users as they readily opt-in to share personal data from their smartphones, Facebook posts, and web browsing activities.

With this ocean of data and personal information to work with, next-generation search developers are positioning their products to engage the user in a process of sense-making in which the search experience moves beyond discovery to contextualization and personalization. Keywords and Boolean searching will become a thing of the past, as searchers use more phrase-like, natural language search strategies, maybe even in the form of voice recognition, to condense multiple variables into one search. The new Facebook graph search exemplifies this process. With one phrase, the Facebook Knowledge Base can be searched with multiple criteria such as people living in your town, who among those people like to ski, and who among *those* "like" a certain ski resort. Imagine phrase-searching in a library's discovery layer to include all articles with a certain author written in the past six months. Of course, we do this now in library catalogs, but one has to know how to do an advanced search and how to use facets and filters to obtain the desired results.

Library knowledge bases will live in the cloud and so too will library management workflows in acquisitions, cataloging, and circulation[10]. In the near future, we will move away from desktop clients to entirely virtual environments where applications are web-based and data is stored in the cloud. OCLC is already making progress toward this end with its Worldshare Management Services[11]. Harnessing the power of the collective, OCLC WMS taps into aggregated bibliographic data in WorldCat, licensed data, vendor records, authority records, and local holdings information. Cataloging, acquisitions, and circulation Workflows that libraries have traditionally managed from desktop clients are now done entirely in a centralized web-based environment, eliminating the need for systems upgrades and maintenance.

In law libraries, the future may follow a different course. If anything, the role of law libraries as intermediaries in the search and discovery realm is becoming increasingly murky as large legal publishers and vendors offer more services native to their own databases. Legal information is becoming more siloed, not less. This model stifles innovations in search and discovery within the broader context of the information economy. Legal information metadata is not exposed to a web-based search engines and the full text itself is locked behind license restrictions. These ideals run contrary to most data models today.

It remains to be seen whether consumers of legal information will drive either a continued dependency on legal information silos or a more open paradigm in which legal information will be incorporated into the larger information economy. The next generation of law students and practitioners will drive these decisions into the next decade.

Where Law Libraries are Headed

Recently, a third year law student asked me to post a list of the ten most frequently used databases to the law library website. As we talked about his request, I let him know that these databases were listed in several areas of our website as well as the website for the main campus library, and were also searchable in the catalog. Nonetheless he wanted just this list, in a place that he thought made sense to him on our website.

Most of us have experienced odd requests from attorneys, law faculty, and law students who demand the easiest possible customizations to their legal research experiences. I often have requests at the reference desk to find news articles that I find so easily with one search in Google that it leads me to worry about the most basic proficiencies of the community I serve.

The generations of legal professionals who have grown up using print will most likely not be on board with the newest technological innovations in legal research technology; however, their successors, our current law students and new lawyers, have a long way to go in becoming proficient in the newest technologies, much less expert users. The development of discovery layer technologies, federated search engines, and web-scale services have all been driven in large part by consumer demand. Yet the legal community remains conservative in its views on making the legal research environment more open to the rest of the information economy. If anything, it seems we've become even more loyal to the largest legal information publishers, who are becoming increasingly proprietary, siloed, and monopolized than ever before. In looking at the quality of these products and the features offered, it's easy to see why. Work folder sharing, litigation profiles, case mapping, and enhanced mobile apps are just a few of the extras offered in Lexis and Westlaw. Relevancy is geared specifically toward the law. Indexes are created by legal experts themselves. And students know they will be using these tools in practice.

Case law research and research in primary sources can be complex and time-consuming. The importance of finding legitimate resources cannot be overstated. Discovery layers and web-based searches are best left to researchers looking for secondary sources, an area of law that is rapidly expanding. Law libraries that are already using Innovative Interfaces products will find the implementation of the Encore discovery layer relatively simple. Academic law libraries, as part of a larger implementation, might want to weigh the costs and benefits of opening their catalog to the campus community at large. If they decide to do so, the costs of implementation and the training of their constituents must be taken into account.

In the end, all of us will need to become more tolerant of the vast array of information resources available in our discovery systems, and we must be open to learning how to navigate

this new terrain. Legal researchers are becoming more interdisciplinary and tech savvy. As the next generations of law students and legal practitioners become part of our organizations and academies, the hope is that these digital natives will drive the further opening-up of the legal information environment.

---

[1] William G. Harrington. "A Brief History of Computer-Assisted Legal Research." *Law Library Journal* 77 (1985): 543–56

[2] Fagan, Jody Condit. "Discovery Tools and Information Literacy." *Journal of Web Librarianship* 5, no. 3 (2011): 171–78.

[3] Shokouhi, Milad, and Luo Si. *Federated Search*. Hanover, MA: now Publishers Inc., 2011.

[4] Vaughan, Jason. "Chapter 1: Web Scale Discovery What and Why?" *Library Technology Reports* 47, no. 1 (2011): 5–11.

[5] Craigle, Valeri. "Discovery Layers in Law Libraries." *AALL Spectrum* 16, no. 3 (2010): 7-9.

[6] Vaughan, Jason. "Investigations into library web scale discovery services." *Information Technology and Libraries* 31, no. 1 (2012): 32–82. http://ejournals.bc.edu/ojs/index.php/ital/article/view/1916/1744

[7] Barr, Jeff, "Amazon Web Services: Building a 'Web-Scale Computing' Architecture." Slideshow. April 2, 2009. http://www.slideshare.net/goodfriday/amazon-web-services-building-a-webscale-computing-architecture.

[8] Craigle, Valeri. "Discovery Layers in Law Libraries." *AALL Spectrum* 16, no. 3 (2010): 7-9.

[9] Breeding, Marshall. "Preparing for the Long-Term Digital Future of Libraries." *Computers in Libraries* 31, no. 1 (2011): 24–26.

[10] Axford, Mary A. "Ultimate Debate Program on Web-Scale Discovery Services. A Report of the LITA Internet Resources and Services Interest Group Meeting, American Library Association Annual Conference, New Orleans, June 2011." *Technical Services Quarterly* 29, no. 2 (2012): 142–147.

[11] Gresham, Anne. "Next-Generation Automation: The Cloud and OCLC's Web-scale Management Services. "http://www.annegresham.com/mls/courses/534/automation/Gresham-Anne-Assignment2.docx".

Bibliography

In General

Arewa, Olufunmilayo. "Open Access in a Closed Universe: Lexis, Westlaw, Law Schools, and the Legal Information Market." *Lewis & Clark Law Review* 10 (2006): 797; Case Legal Studies Research Paper No. 06-03; Northwestern Public Law Research Paper No. 07-02. Available at SSRN: http://ssrn.com/abstract=888321.

Breeding, Marshall. "Preparing for the Long-Term Digital Future of Libraries." *Computers in Libraries* 31, no. 1 (2011): 24–26.

Childers, Scott. "An Intoduction to Relevance-Ranking Systems." *Faculty Publications, UNL Libraries*. Paper 117 (2006). http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1123&context=libraryscience.

Hoeppner, Athena. "The Ins and Outs of Evaluating Web-Scale Discovery Services." *Computer in Libraries* 32, no. 3 (2012)

Fagan, Jody Condit. "Discovery Tools and Information Literacy." *Journal of Web Librarianship* 5, no. 3 (2011): 171–78.

OCLC WorldShare Management Services website. http://www.oclc.org/webscale/default.htm.

Singhal, Amit. "Giving You Fresher, More Recent Search Results." *Google Official Blog.* November 3, 2011. http://googleblog.blogspot.com/2011/11/giving-you-fresher-more-recent-search.html.

Weare, William H., Jr., Sue Toms, and Marshall Breeding. "Moving Forward: The Next-Gen Catalog and the New Discovery Tools." *Library Media Connection* 30, no. 3 (2011): 54–57.

Westrum, Anne-Lena. "The Key to the Future of the Library Catalog is Openness." *Computers in Libraries* 31, no. 3 (2011): 10–14.

Web-Scale Discovery

Axford, Mary A. "Ultimate Debate Program on Web-Scale Discovery Services. A Report of the LITA Internet Resources and Services Interest Group Meeting, American Library Association Annual Conference, New Orleans, June 2011." *Technical Services Quarterly* 29, no. 2 (2012): 142–147.

Barr, Jeff, "Amazon Web Services: Building a 'Web-Scale Computing' Architecture." Slideshow. April 2, 2009. http://www.slideshare.net/goodfriday/amazon-web-services-building-a-webscale-computing-architecture.

Elder, Carol, et al. "Single Search? The Promises and Pitfalls of Web-Scale Discovery." *BCLA Browser: Linking the Library Landscape* 3 no. 3 (2011).

Gross, Julia, Lutie Sheridan. "Web Scale Discovery: The User Experience." *New Library World* 112 no. 5/6 (2011): 236–47. doi:10.1108/03074801111136275.

Meadow, Kelly, and James Meadow. "Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis." *College & Undergraduate Libraries* 19, no. 2-4, (2012): 163–75.

Howard, David, and Constance Wiebrands. "Culture Shock: Librarians' Response to Web Scale Search." *Edith Cowan University Research Online* (2011). http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=7208&context=ecuworks.

Vaughan, Jason. "Investigations into library web scale discovery services." *Information Technology and Libraries* 31, no. 1 (2012): 32–82. http://ejournals.bc.edu/ojs/index.php/ital/article/view/1916/1744

Federated Search

Shokouhi, Milad, and Luo Si. *Federated Search*. Hanover, MA: now Publishers Inc., 2011.

Discovery Layers

Becher, Melissa, and Kari Schmidt. "Taking Discovery Systems for a Test Drive." *Journal of Web Librarianship* 5, no. 3 (2011): 199–219.

Caplan, Priscilla. "On Discovery Tools, OPACs and the Motion of Library Language." *Library Hi Tech* 30, no. 1 (2012): 108–15.

Craigle, Valeri. "Discovery Layers in Law Libraries." *AALL Spectrum* 16, no. 3 (2010): 7-9.

Little, Geoffrey. "Thinking about Discovery Layers." *The Journal of Academic Librarianship* 38, no. 6 (2012): 346–47.

Vaughan, Jason. "Chapter 1: Web Scale Discovery What and Why?" *Library Technology Reports* 47, no. 1 (2011): 5–11.

———. *Web Scale Discovery Services*. *Library Technology Reports* 47, no. 1 (2011). Chicago: American Library Association, 2011. PDF e-book.

Yang, Sharon Q., and Melissa A. Hofmann. "Next Generation or Current Generation?: A Study of the OPACs of 260 Academic Libraries in the USA and Canada." *Library Hi Tech* 29, no. 2 (2011): 266–300.

The Cloud

Hartman, Robin R. "Life in the Cloud: A WorldShare Management Services Case Study." *Journal of Web Librarianship* 6, no. 3 (2012): 176–85.

Gresham, Anne. "Next-Generation Automation: The Cloud and OCLC's Web-scale Management Services.   "http://www.annegresham.com/mls/courses/534/automation/Gresham-Anne-Assignment2.docx".

Wale, Carla P. "Cloudy with a Chance of Open Source: Open Source Integrated Library Systems and Cloud Computing in Academic Law Libraries." *Legal Reference Services Quarterly* 30, no. 4 (2011): 310–31.

History

William G. Harrington. "A Brief History of Computer-Assisted Legal Research." *Law Library Journal* 77 (1985): 543–56