



Managing Search Engine Optimization: An Introduction for Library Administrators

Kenning Arlitsch, Patrick OBrien, and Brian Rossmann

This is a preprint of an article that originally appeared in the Journal of Library Administration in 2013.

Kenning Arlitsch , Patrick OBrien & Brian Rossmann (2013) Managing Search Engine Optimization: An Introduction for Library Administrators, Journal of Library Administration, 53:2-3, 177-188
<http://dx.doi.org/10.1080/01930826.2013.853499>

Made available through Montana State University's [ScholarWorks](http://scholarworks.montana.edu)
scholarworks.montana.edu

Column Title: posIT

Column Editor: Kenning Arlitsch, Dean of the Library, Montana State University, Bozeman, MT
kenning.arlitsch@montana.edu

This JLA column posits that academic libraries and their services are dominated by information technologies, and that the success of librarians and professional staff is contingent on their ability to thrive in this technology-rich environment. The column will appear in odd-numbered issues of the journal, and will delve into all aspects of library-related information technologies and knowledge management used to connect users to information resources, including data preparation, discovery, delivery and preservation. Prospective authors are invited to submit articles for this column to the editor at kenning.arlitsch@montana.edu

Managing Search Engine Optimization: An Introduction for Library Administrators

by Kenning Arlitsch, Patrick OBrien, and Brian Rossmann

Introduction

Libraries collectively spend millions of dollars each year creating websites and digital repositories, but optimizing for search engines is too often an afterthought and makes digital library use a fraction of what it could be. Even libraries that do take search engine optimization (SEO) into account tend to relegate its practice to a few individuals in IT departments, often resulting in a disjointed and unproductive program that is viewed as a limited domain rather than a primary concern of the entire organization. SEO deserves cabinet-level attention because of its potential to help libraries reach more users and derive assessment data. This article is aimed at giving library administrators a high-level perspective of SEO so that they may be equipped to ask the right questions of their technical staff, software vendors and content suppliers. It stresses the importance of aligning SEO with institutional priorities and integrating it into the strategic plan. SEO is

most effective when it is an organizational priority and when it is understood and driven by administrative teams.

Libraries and their funders make “significant investments in the digitization” (Maron & Pickle, 2013) of special and other collections each year. The entire scope of a digital library includes far more than scanning objects and loading them into a database, however, and the cost of developing and sustaining a digital library requires substantial funding. Whether costs are borne by internal sources or from external grants and donations, funding providers have become more interested in the value and use of their investments.

Libraries would do well to emphasize Internet search engines over their own websites as a means of bringing users to digital libraries. The concept of “inside-out” library resources (Dempsey, 2010) emphasizes the value of search-engine discoverability, overriding the more traditional “outside-in” expectation that users will begin research in the library. Americans submit approximately 20 billion search queries to Google each month, and Google represents “only” 66% of the domestic search engine market share (comScore, 2013). Most library websites simply do not draw much direct traffic of their own (DeRosa et al., 2010) and it is a much better investment of time and energy to establish good relationships with search engines and social media sites than it is to try to draw users directly to library websites. Establishing good relationships means ensuring that search engine crawlers can navigate through websites and repositories without running into design, metadata and systems barriers that negatively impact index ratios or rankings in search engine results pages (SERP). Search engines must find machine-readable and comprehensible text to index, and they must be convinced that their

customers will have a worthwhile experience at a library site. Slow websites, over-use of graphics, dead links, and poor-quality content or redundant metadata are factors that may contribute to a search engine's decision to send its customers elsewhere.

Some may dismiss the practice of SEO as “gaming” search engines, but it's only gaming if it is done poorly and with malicious intent. Practiced according to the guidelines and tools provided by the major search engines (Microsoft, Inc., 2013) (Google, 2013), SEO helps libraries establish better relationships with search engines, and thus with the customers of those search engines. Libraries gain users, perhaps many thousands of users, when search engines decide to send their customers to library sites. Previous research by the authors at the University of Utah's J. Willard Marriott Library produced a 500% increase in referrals from Google that resulted in a 132% increase in visits to digital collections (Arlitsch & OBrien, 2013) by implementing good SEO practices.

SEO Across the Organization

Search engines index digital files and their metadata, and therefore SEO is naturally considered to be the concern of people who manage technology. While it is true that effective SEO affects the layers of technology that manage and deliver digital content, it really is the concern of anyone whose work is represented on the Internet because it impacts how accessible that work is to the intended audiences. SEO is not something that should be left exclusively to an organization's IT department: “an IT department should not be left to make, often by default, the choices that determine the impact of IT on a company's business strategy”(Ross & Weill, 2002). Nearly every library employee has some interest or influence in the content and the services that a library makes available on

the Web, and that makes for many stakeholders. Both technology and people should be driven by the strategic goals of the organization.

Strategic Planning

Just as it takes many people to make print collections accessible to the public, SEO practiced broadly will help ensure that items in the digital library will be accessible as well. Some SEO solutions are technical in nature, but all aspects of SEO require communication, management, and coordination to ensure that people are working together to achieve common goals. Including SEO goals and objectives in a library's strategic plan elevates its importance and creates a better chance that SEO techniques will be applied effectively and its results measured accurately. Driving SEO strategically will help ensure its success, and will set goals and objectives that allow administrators to gather accurate use statistics so that reporting becomes an integral part of the assessment process. Data gathered from SEO analysis tools can help make course corrections, communicate a narrative about what the library is trying to achieve for its users, and enhance library assessment.

Strategic alignment with the institution's mission

The success rate of a library's SEO efforts improve dramatically when they are aligned with the strategic goals of the institution. A carefully communicated and well-executed institutional repository (IR) strategy can deliver value recognized by university administrators and its faculty. For example:

1. Create a showcase for research and scholarship produced at the institution that can be leveraged for fundraising purposes with legislatures, alumni, granting agencies, and local communities.
2. Increase faculty citation rates by making publications and data sets easily accessible to a wider audience. (Piwowar, Day, & Fridsma, 2007). Increased citation rates could, in turn, raise university rankings (The Times Higher Education, 2010).
3. Provide a hedge against publisher price inflation that continues to exceed the Consumer Price index (Bosch, Henderson, & Klusendorf, 2013).

Framing the value this way at Montana State University has resulted in the Library's IR efforts being elevated to a line item in the University's academic strategic plan and has in turn helped to drive part of the Library's strategic plan (see Figure 1). It is also leading to easier conversations with MSU faculty about why they should deposit their papers in the University's IR and why faculty should consider adopting an open access publishing policy.

[Insert Figure 1]

A well-executed IR strategy produces content that can be found by search engines, and in particular by academic engines like Google Scholar and Microsoft's Academic Search. These search engines will only index metadata that they can parse so that they can deliver citations in whatever styles their users require, and that means citation data in IR's must be offered in discrete fields. Earlier research has demonstrated that library-

developed IR's often fail in this regard because they tend to lump citation data into single fields, which is not machine-comprehensible (Arlitsch & O'Brien, 2012).

Libraries can use SEO strategy to think about adding value beyond their traditional roles and boundaries. For example, using knowledge derived from SEO research they can help develop the metadata schema used to store citation information in a campus faculty activity database used for evaluations of faculty performance. Engaging at this level allows libraries to influence the alignment between the schema employed in faculty performance reviews, in promotion and tenure reviews and the IR metadata. This alignment can increase faculty participation by incentivizing them to deposit published research, reduce effort needed to populate IR citation metadata, improve IR data quality, and ensure scholarly output is discoverable by the search engines used by academic researchers.

Library administrators must have metrics defined and tools in place to evaluate the results of their SEO efforts and objectively demonstrate the value they are delivering to their stakeholders. The first step is to determine what stakeholders value and what can be objectively measured. The best metric (within the technical reach of most libraries) for evaluating the value of an IR is the number of IR PDF views or downloads by a target audience. Visitors being referred to the IR by Google Scholar, or who have an IP address that resolves to a .edu domain, are more likely to be academics than typical visitors from Facebook or Yahoo!. Academic visitors are more likely to formally cite papers they choose to view or download as PDF files. It is worth paying attention to SEO areas that will help increase faculty citation rates.

Setting Expectations and Using Metrics

Statistics have long been a part of our business. Libraries have always counted the size of their collections—the number of items borrowed, interlibrary loans, expenditures—and they have always used those numbers to create comparisons to other libraries. SEO continues this tradition of measurement, with additional complexity and some variation in the vocabulary. Gate counts for physical visits are still needed, but counting online visitors is at least as important, and can actually tell administrators a whole lot more about their users (in anonymous terms) than traditional counts.

Libraries must move beyond counting basic page views and visits to individual websites. They must break down silos and look at the entire user click stream across all their web properties to create metrics their internal and external stakeholders value. This is an important concept: it is not enough to track user behavior on a single web server because that provides only a partial picture. One physical server may contain the library's website, another server hosts a digital repository, and a third may serve streaming video to users, but a user who peruses a particular digital collection may touch all three servers in a visit. A single machine can also host multiple web servers with distinct URLs established to promote a featured collection or program. While these servers are probably all in the organization's domain, reporting tools must be set up appropriately to track users as they move from one to another. (see the section on “Traditional SEO Components”)

A dashboard for each digital collection could help the library's leadership team have discussions with current and future donors, as well as help collection managers who want to demonstrate value to grant providers and collaboration partners. Dashboard metrics

can provide objective data that support a dialogue with the university's administration about the value of the IR and how individual faculty members, departments, and colleges can better utilize it.

How Search Engines Find Websites

Most websites and digital repositories are driven by databases; very few are still collections of static HTML pages. Some databases are relational, meaning that the data reside in table structures, and some are flat text-based files that incorporate markup like XML to provide structure for the data. In either case, web pages from database-driven sites are constructed dynamically, or “on the fly.” A programmed script assembles various components from the database to generate a page at the moment that it is “called” by the user who clicks on a hyperlink or types a URL in a web browser. Data components may include the object itself (a digitized photograph, for example), the metadata associated with it, and the HTML template that will include a header and footer along with other design elements that help make a page useful and attractive.

Search engines find, harvest, and index websites through the use of programs called “crawlers,” “robots,” or “spiders.” A search engine crawler does not actually crawl through a database. It wants to see the compiled page just as the user sees it, so it follows the link for each object in the database and triggers the generation of a page for each. It then harvests the text that is generated for that page before moving on to the next link. It is at this crucial juncture, when the page is displayed, that all the text a library hopes will be indexed by the search engine must be present, prioritized, and accessible to the crawler. It's crucial to note that search engine crawlers are not unlike visually impaired

users in that they “can’t read text in images, can’t interpret JavaScript or applets, and can’t ‘view’ many other kinds of multimedia content” (Hagens, 2005).

Traditional SEO Components

We distinguish “traditional” SEO from more recent “semantic” SEO developments because the traditional components form a foundational relationship that allows search engines to find, harvest, and index websites and digital repositories. Without that foundation any additional value provided by semantic web techniques is marginal at best, and irrelevant in most cases. Semantic SEO does hold enormous potential for bringing more accurate and relevant search results to the user as it helps set context and meaning for search engines, but this topic is outside the scope of this paper.

Traditional SEO involves identifying the requirements of search engines, establishing communication channels to assure goals are being met, and monitoring the relationship for disruptive changes that can occur on either side. The basic goals of SEO are to have web content included in a search engine’s index (indexing ratio) and to rise to the top of SERP (search engine results page), otherwise known as “rank” for searches conducted by target audiences. Specific traditional SEO actions that help create that foundational relationship with search engines include:

1. Develop internal inventories of the organization’s logical and physical domains.
 - a. Logical domains are the domains, sub-domains, and sub-directories used to organize Web content and are typically aligned with the people responsible for the content in a given collection or repository.

- b. Physical domains describe the physical servers that host the digital library's logical domains. A server may host multiple logical domains, such as an IR, a departmental website, or a digitized collection.
2. Designing websites and navigational paths that pose no barriers to crawlers. Graphics can make a website attractive, but their overuse tends to increase page load time and provides little or no text for crawlers to index. Complex internal link structures or labyrinthine paths to objects reduce crawler efficiency and limit the total amount of content they can find and index on the site. Crawler efficiency can weigh heavily in how a website and its content is ranked in SERPs.
3. Serving indexable text and metadata. Search engines do not tolerate placement of invisible text or keyword stuffing, and redundant or repetitive metadata lead crawlers to think that the objects being described are identical.
4. Configuring servers and software to deliver results quickly to users and search engine crawlers. Slow server response, dead links, and failure to communicate location changes and downtime through an accepted set of messages is viewed as a bad user experience.
5. Setting up Webmaster Tools and analytics software for monitoring and assessment purposes. Webmaster Tools provide the feedback loop from search engines as they try to crawl a library's sites and repositories, letting them know what kind of problems they encounter. Analytics software such as Google Analytics provides a wealth of information about a library's visitors.

Assessment

Accountability has been important in higher education for many decades (Marrs, 2009). Now, however, in times of curtailing budgets and sparse resources, colleges and universities are being more closely scrutinized than ever. They are being expected to demonstrate that they are managing their budgets wisely, provide good value to students and faculty, and have a positive impact on student success. Accordingly, libraries are increasingly developing cultures of assessment (Farkas, 2013). “Not only does assessment give librarians a venue for communicating with stakeholders, it determines “the fit” between institutional mission and achieved outcomes, articulates effectiveness, fosters improvement, increases efficiency, and demonstrates accountability” (Oakleaf, 2010). As libraries invest more of their resources creating a digital presence, it is important that the findability and use of digital collections becomes a part of libraries’ assessment efforts so that they can demonstrate the value of their digital presence. “Institutional assessment efforts should not be concerned about valuing what can be measured, but instead about measuring what is valued.” (Banta, 1996). Assessment librarians and administrators involved in implementing and monitoring SEO can gather more accurate data on the use of digital collections. That SEO data then becomes another metric on the library’s scorecard in the assessment cycle.

In addition to incorporating SEO into a library’s assessment cycle, it is important to continually assess SEO activities themselves. Assessment librarians, if they are provided with access to Webmaster Tools, and if they incorporate SEO assessment into their workflow, can monitor changes in the usage patterns of digital collections. When anomalies are observed, they may engage administrators, digital initiatives librarians, and

IT staff to identify solutions. “Assessment “strives to know ... what is” and then uses that information to change the status quo” (Keeling & International Center for Student Success and Institutional Accountability, 2008).

Incorporating SEO into a library’s assessment program will pay dividends both with respect to improving SEO itself and allowing libraries to measure and demonstrate increased value to library stakeholders.

Webmaster Tools

Webmaster Tools are offered both by Google and Bing to assist technical teams in identifying and addressing issues that will help sites perform better in search results. The search engines’ crawlers report the following information and more about each site that is verified via the Webmaster Tools product:

- Identify which parts of the site pose problems for crawlers
- Notify the search engine of new or revised XML sitemaps
- Generate and analyze the robots.txt files
- Remove URLs from the crawl when they no longer exist
- Identify issues with page titles and meta tags
- Identify the top search terms used to reach sites
- Review pages as the search engine crawler would see them
- Provide notifications of any quality guideline violations
- Provide statistics

Google Analytics

Numerous commercial and free website analysis software tools exist. Some of them analyze web server logs, while others utilize page tagging techniques that embed code into each HTML page of a website to set and track “cookies.” With page tagging, the code sends a message to a third party system each time a page is viewed, and it compiles visitor information concerning sessions, page views and traffic sources (e.g., referring sites, search terms, etc.). There are advantages and disadvantages to both methods. Google Analytics (GA) utilizes page tagging and is suggested for its ease of use, zero cost, excellent support, and power. If configured properly, GA can provide data about a library’s physical domain that will help administrators understand where visitors are coming from and what they are looking for. Aside from creating a Google Account, configuring the GA product and embedding a bit of code in each web page HTML header, there is no further overhead for basic reporting.

Google Analytics provides powerful and anonymous data about visitors to websites, information about their behavior while they are visiting, and identifies the tools they use to view the site. It can also help to troubleshoot quickly problems on the site. Administrators may learn that certain page titles are inaccurate or could be written more descriptively, or that some users have bookmarked and are still visiting an obsolete page that has never been deleted from the server. GA can provide the following pieces of information, among others, about a website and its visitors:

- Number of unique and returning visitors
- Search terms used to reach a specific page
- Operating systems and browsers used by visitors

- Whether a mobile device was used for the visit and, if so, what type
- What pages were landed on and from which ones users exited
- Most viewed pages, the order in which they were viewed, and for how long
- The countries and cities from which searches originate
- How long pages take to load (as mentioned earlier, slow-loading pages can cause search engines to send their customers elsewhere)

Google Analytics can even establish and track goals. Administrators might want to see which academic papers were downloaded from the institutional repository, and if configured properly, GA can provide results by college, department, and author. This can help libraries inform faculty about the frequency with which their papers are accessed, and perhaps will help generate support for increased faculty IR participation.

Cross-domain tracking helps track users when they are referred from one logical domain to another. For many digital repositories, the information describing a collection is contained in one logical domain (e.g., a website located at lib.montana.edu), while the actual objects in the collection are located within a different logical domain (e.g., an IR located at scholarworks.montana.edu). Without cross-domain tracking set up, all that can be known is that visitors were referred from one logical domain to the other. The critical information about the links and search terms used that led to viewing or downloading the object within the IR will not be passed from the visitors' entry point. In other words, administrators won't know why visitors were referred to the IR, or how they got there. Cooperation among domains will help identify and influence key elements concerning visibility of the library's content within SERPs.

The information that Google Analytics provides creates a feedback loop that can help repository managers improve the user experience, and to identify what is working well and what is not. It can help improve the text describing a given page, resulting in a better fit with the traffic sources and search queries that deliver users to the library's sites.

Summary

Digital libraries suffer from lack of visitation and use because few libraries are proactive and strategic about search engine optimization. A well-executed SEO strategy will connect users to the information they seek and can play a significant role in increasing citation rates of academic research. Library administrators can derive valuable data about their organizations through a carefully managed SEO program that ensures websites and digital object metadata are harvestable and comprehensible by search engine crawlers. SEO affects many areas of an organization, and in turn there are numerous people who play roles in its successful practice. Those roles are best driven from a strategic plan that aligns with institutional goals and is driven by the library's leadership. A wealth of free software exists to help troubleshoot problems that search engine crawlers encounter when they try to harvest websites and digital repositories, and Web analytics software like Google Analytics can provide rich data about visitor behavior that can help organizations make course corrections for better service.

References

- Arlitsch, K., & O'Brien, P. S. (2012). Invisible institutional repositories: Addressing the low indexing ratios of IRs in Google Scholar. *Library Hi Tech*, 30(1), 60–81.
doi:10.1108/07378831211213210

- Arlitsch, K., & OBrien, P. S. (2013). *Improving the visibility and use of digital repositories through SEO*. Chicago: ALA TechSource, an imprint of the American Library Association. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=578551>
- Banta, T. W. (1996). *Assessment in practice: putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Bosch, S., Henderson, K., & Klusendorf, H. (2013, April 25). The Winds of Change: Periodical Price Survey 2013. *Library Journal*. Retrieved August 11, 2013, from <http://lj.libraryjournal.com/2013/04/publishing/the-winds-of-change-periodicals-price-survey-2013/>
- comScore. (2013, July 12). comScore Releases June 2013 U.S. Search Engine Rankings. *comScore, Inc*. Retrieved July 12, 2013, from http://www.comscore.com/Insights/Press_Releases/2013/7/comScore_Releases_June_2013_U.S._Search_Engine_Rankings
- Dempsey, L. (2010, January 11). Outside-in and inside-out. *Lorcan Dempsey's weblog on libraries, services and networks*. Retrieved from <http://orweblog.oclc.org/archives/002047.html>
- DeRosa, C., Cantrell, J., Carlson, M., Gallagher, P., Hawk, J., & Sturtz, C. (2010). *Perceptions of Libraries, 2010: Context and Community* (p. 108). OCLC, Inc. Retrieved from <http://www.oclc.org/reports/2010perceptions.htm>

- Farkas, M. G. (2013). Building and sustaining a culture of assessment: best practices for change leadership. *Reference Services Review*, 41(1), 13–31.
doi:10.1108/00907321311300857
- Google. (2013, July 17). Webmasters – Google. Retrieved August 17, 2013, from <https://www.google.com/webmasters/>
- Hagans, A. (2005, November 8). High Accessibility Is Effective Search Engine Optimization. *A List Apart*, 207. Retrieved from <http://www.alistapart.com/articles/accessibilityseo>
- Keeling, R. P., & International Center for Student Success and Institutional Accountability. (2008). *Assessment reconsidered: institutional effectiveness for student success*. [United States]: ICSSIA.
- Maron, N. L., & Pickle, S. (2013, February 21). Appraising our digital investment: sustainability of digital special collections in libraries. Association of Research Libraries and Ithaka S+R. Retrieved from <http://www.arl.org/storage/documents/publications/digitizing-special-collections-report-21feb13.pdf>
- Marrs, H. (2009). Perceptions of College Faculty Regarding Outcomes Assessment, 13(2). *IEJLL: International Electronic Journal for Leadership in Learning*, 13(0). Retrieved from <http://iejll.synergiesprairies.ca/iejll/index.php/iejll/article/view/688>
- Microsoft, Inc. (2013, July 17). Webmaster Guidelines - Bing Webmaster Tools. Retrieved August 17, 2013, from <http://www.bing.com/webmaster/help/webmaster-guidelines-30fba23a>

- Montana State University. (2013). MSU Library Strategic Plan 2013. Montana State University Library. Retrieved from http://www.lib.montana.edu/about/msu_library_strategic_plan.pdf
- Oakleaf, M. (2010). The value of academic libraries: a comprehensive research review and report. Association of College and Research Libraries. Retrieved from <http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/value>
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*, 2(3), e308.
doi:10.1371/journal.pone.0000308
- Ross, J. W., & Weill, P. (2002). Six IT decisions your IT people shouldn't make. *Harvard Business Review*, 80(11), 84–95.
- The Times Higher Education. (2010). The Times Higher Education World University Rankings 2010-2011. Retrieved October 4, 2011, from <http://www.timeshighereducation.co.uk/world-university-rankings/>

Figure 1: Strategic Plan example (Montana State University, 2013)

Ensure staff and faculty understand the strategic importance of SEO efforts

Montana State University Strategic Plan

Objective: Enhance infrastructure in support of research, discovery and creative activities

University Metric: MSU will increase grant-sponsored investment in centers, core facilities and resources to expand state of the art tools, expertise, and opportunities for research and creative activities

- Develop institutional repository of intellectual output of campus

MSU Library Strategic Plan

Objective: Improve discovery layers and findability of library content

SEO Program Activities

Library Metric: Increase indexing rates of MSU content to improve access for researchers worldwide

- Create dashboard for analytical reporting
- Establish a baseline for Google's indexing of Library web content for search engine optimization
- Establish baseline for Google Scholar's indexing of ScholarWorks