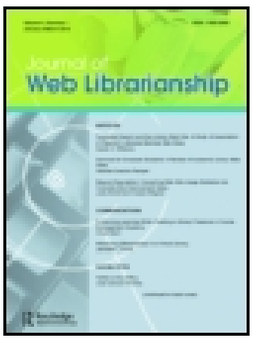

Journal of Web Librarianship

ISSN: 1932-2909 (Print) 1932-2917 (Online) Journal homepage: http://www.tandfonline.com/loi/wjwl20

The Open SESMO (Search Engine & Social Media Optimization) Project: Linked and Structured Data for Library Subscription Databases to Enable Web-scale Discovery in Search Engines

Jason A. Clark & Doralyn Rossmann

To cite this article: Jason A. Clark & Doralyn Rossmann (2017): The Open SESMO (Search Engine & Social Media Optimization) Project: Linked and Structured Data for Library Subscription Databases to Enable Web-scale Discovery in Search Engines, Journal of Web Librarianship, DOI: 10.1080/19322909.2017.1378148

To link to this article: http://dx.doi.org/10.1080/19322909.2017.1378148

JOURNAL OF WEB LIBRARIANSHIP
https://doi.org/10.1080/19322909.2017.1378148

Routledge
Taylor & Francis Group

The Open SESMO (Search Engine & Social Media Optimization) Project: Linked and Structured Data for Library Subscription Databases to Enable Web-scale Discovery in Search Engines

Jason A. Clark and Doralyn Rossmann

Montana State University Libraries, Bozeman, Montana, USA

ABSTRACT

Today's learners operate in digital environments which can be largely navigated with no human intervention. At the same time, libraries spend millions and millions of dollars to provide access to content which our users may never know is available to them. Through the Open SESMO (Search Engine & Social Media Optimization) database project, Montana State University (MSU) Library applied search engine optimization and structured data with the Schema.org vocabulary, linked data models and practices, and social media optimization techniques to all the library's subscribed databases. Our research shows that Open SESMO creates significant return-on-investment with substantial increased traffic to our paid resources by our users as evidenced through analytics and metrics. In the core research of the article, we take a quantitative look at the pre/post results to assess the Open SESMO method and its impact on organic search referrals and use of the collection analyzing data from three distinct fall semesters. Returns include demonstrated library value through database recommendations, connecting researchers to subject librarians, and increased visitation to our library's paid databases with growth in organic search referrals, impressions, and click-through rates. This project offers a standard and innovative practice for other libraries to employ in surfacing their paid databases to users through the open web by applying structured and linked data methods.

ARTICLE HISTORY
Received 3 April 2017
Accepted 6 September 2017

KEYWORDS
electronic resources; HTML; linked data; RDFa; Schema.org; search engine optimization (SEO); social media optimization (SMO); structured data

Introduction

Today's learners operate in digital environments which can be largely navigated with no human intervention. At the same time, libraries spend millions and millions of dollars to provide access to content which our users may never know is available to them. Through the Open SESMO (Search Engine & Social Media

CONTACT Digital Initiatives Librarian Jason A. Clark jaclark@montana.edu Montana State University Libraries, PO BOX 173320, Bozeman, MT 59717-3320, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wjwl.

Optimization) database project, Montana State University (MSU) Library applied search engine optimization and structured data with the Schema.org vocabulary, linked data models and practices, and social media optimization techniques to all the library's subscribed databases (e.g., Academic Search Complete, Web of Knowledge, PsycInfo). At its core, Open SESMO provides a model of best practices for library collections markup, creates metadata routines for web-scale external vocabularies, and measures the impact of applying optimizations and metadata to library collections. More specifically, we show in our research that Open SESMO creates significant return-on-investment with substantial increased traffic to our paid resources by our users as evidenced through Google Analytics, Facebook Analytics, and Twitter Analytics. Underlying this project is the concept of meeting users where they are (i.e., on search engines and in social media networks) as advocated in the Library 3.0 concept. This project offers a standard and innovative practice for other libraries to employ in surfacing their paid databases to users through the open web by applying structured and linked data methods.

Like many libraries, we have given much attention to making our homegrown digital collections discoverable by search engines and optimized for sharing on social media networks. Similarly, we have spent much time and effort making our paid subscriptions optimized for teachable moments through our website and through our library guides hosted by Springshare. In this study, we offer an example of taking lessons from the presentation of both locally digitized collections and paid collections, and applying these lessons holistically to our paid databases, making them openly discoverable by our users through Internet search engines and social media networks, treating them as any of our other digital collections.

We have based our research on several studies that reinforced our interest in meeting our users where they are. A study from OCLC in 2011 noted that "search engines continue to dominate, topping the list of electronic sources most used to find online content (93 percent), followed closely by Wikipedia (88 percent). The key difference in usage between search engines and Wikipedia is the frequency of use 75 percent of students who use search engines do so daily, compared to 20 percent of those who use Wikipedia" (Gauder 2011). Roger Schonfeld's (2015) Ithaka S+R report, "Meeting Researchers Where They Start," pointed out that the "library is not the starting point" and even further that "the proxy server is not the answer." This project meets our scholars where they start (in commercial search engines, like Google) and increases the likelihood that they will discover, access, and share our subscribed resources in search engines and social media networks.

Open SESMO methods are based on several components: structured semantic markup, search engine and social media optimization, and employing linked data techniques. Structured semantic markup is the act of providing rich descriptions of the "aboutness" of web pages by embedding additional tags into HTML pages or constructing alternate data formats like JSON-LD to carry these enhanced descriptions. In this case, Search Engine Optimization (SEO) includes a restructuring of the site architecture, creating a sitemap to list the inventory of pages for

indexing, and the assigning of canonical URLs to each page to identify the preferred URL to appear in Search Engine Result Pages (SERPs). Social Media Optimization (SMO) includes adding Facebook Open Graph Tags and Twitter Card markup to database pages, as described on the Twitter Developer and Facebook Developer sites. Finally, the linked data techniques applied during the project were based on the external enumerations specification and consisted of linking to external linked open data vocabularies to connect page descriptions to the linked open data cloud.

In this article, we provide a snapshot of Open SESMO and its application, including steps involved in presenting library-subscribed databases openly as findable, machine-readable digital collections, the methods of promotion for these items, and the ways in which this activity reaches new and disintermediated researchers. In the core research of the article, we take a quantitative look at the pre- and post-results to assess the Open SESMO method and its impact on organic search referrals and use of the collection analyzing data from three distinct fall semesters. These lessons have broad adaptability to any library seeking to connect commercial resources to today's learners in a digital environment.

Literature review

The methods and techniques that became part of this project required a synthesis of several research threads. There just was not a single article or study that touched on all the components. To this end, we are shaping this literature review under the following threads and noting how and why each thread informed the project.

Library 3.0

Well-established models of Web 2.0 and Library 2.0 focus on meeting users where they are, such as blogging sites, and promoting user-generated content, such as library websites that allow patron-created book reviews. Noh (2015) extended these models to Web 3.0 and Library 3.0, in which the library meets the users where they are by generating content that is understandable by machines so that users can find the content through those machines. Likewise, Kwanya, Stillwell, and Underwood (2014) advocated that libraries engage in semantic web technologies to connect users with resources. With the Library 3.0 model in mind, Open SESMO takes content that has been traditionally designed for mediated use through library instruction and browsing, and makes them discoverable through search engines and social networks.

Promoting paid electronic collections

The literature suggests that librarians routinely seek ways to promote their paid collections through marketing and outreach. Some libraries do this to prove their value to patrons. Monie and Clark (2013, 32) observed that "the library needed to

approach our customers with resources that can actually 'help and supply' rather than waiting for customers to 'need' us." They employed an electronic Table of Contents Service as one way to promote these resources. Other libraries promote paid library resources to users to make sure they are getting access to quality resources paid through their tax or tuition dollars, and they continue to have access to these subscriptions because of high use. They focus on one example such promotion by highlighting a database-of-the-week to faculty through a weekly e-mail (Dugan 2011).

Alford (2009) outlined other potential components of a marketing plan for e-resources. In preparing a plan, she recommended, "You need to put some time in to determine what your users think, want, and need" (273). Through focus groups or interviews, your library may determine that a video promotion, web banners on your library's homepage, posters, or other means of outreach may meet users where they are in a way that resonates with your community. At our library, we were already promoting e-resources through means such as library instruction, display monitors in our library cafe, and LibGuides, but Alford's article suggested that we could be present in more spaces where our users are such as search engines and social media networks.

Several studies highlight the use of social media to promote electronic collections. Gustafson, Sharrow, and Short (2017) included a discussion about how to discover the social media networks in use by your specific community and how the library can promote the library, including e-resources, in these networks. Vassilakaki and Garoufallou (2014) explored the literature regarding libraries using Facebook as a means of marketing e-resources, including advertising, creating groups, promoting posts, posting to library employee personal Facebook pages, and posting to the library's Facebook page. Fry (2014) outlined a database promotion campaign and the importance of reviewing usage statistics to determine the success of the effort. While these articles highlighted social media as a place where people get information, they did not address how to optimize our resources for social media sharing.

Each of these articles inspires active promotion of paid library databases as part of a marketing strategy. The Open SESMO project takes these ideas a step further by exploring how optimize the library's website to meet the library community through search engines and social networks to find the paid resources provided by the library.

Promoting digital collections

The literature provides ideas for promoting locally-grown digital collections to users that helped inspire the Open SESMO project. The University of North Texas Library actively reviewed Wikipedia entries that aligned with their digital collections and edited these pages to include links back to UNT Digital Collections (Belden 2007). This proactive effort to be in Wikipedia aligned with our idea of

meeting users where they are. Additionally, the metadata used to describe digital collections both within library systems and for search engines can influence discoverability including schema and models such as FRBR, MARC, RDFa, Linked Data, and Dublin Core (Solodovnik 2011). Institutional repositories have received attention for search engine discoverability through semantic markup of HTML content. These techniques can be applied to the library's paid databases for discoverability by users (Hilliker, Wacker, and Nurnberger 2013).

SEO and semantic identity

Much of the inspiration for the Open SESMO work stemmed from an earlier project in which the authors participated which focused on the library as an entity and how it is understood by web indexing systems. In an article co-written by the present article's authors, Arlitsch and colleagues (2014) encouraged the library community to think about the library's semantic identity and how its content appears to search engines. SEO is often an afterthought for libraries, while it could be used much more actively to help connect users with more library resources (Arlitsch, O'Brien, and Rossmann 2013). The novelty of the present article is that it takes the concepts described in the SEO articles, combines them with the SMO techniques (described in the next section), and applies them to a library setting to paid library databases. Results from this study offer insights for future research and suggest approaches other libraries may adopt by increasing web and social media understanding, and discoverability of their resources.

SMO

In 2006, Rohit Barghava built on the concept of SEO and coined the term "Social Media Optimization." He later updated these rules once his theory had been more widely tested in its applicability to social media networks (2010). In 2016, one of this article's authors described how to adapt the use of SMO idea in libraries (Rossmann and Young). As noted, "Social media optimization (SMO) is a programmatic strategy for building and engaging community" (Rossmann and Young 2016, 5). Part of the SMO strategies include "creating shareable content" and "proactively share." These strategies encourage adding web page coding to control how pages appear when shared via social media, putting social share buttons on web pages, and actively sharing content on social networks, which we adapted to the Open SESMO project. Onaifo and Rasmussen (2013) advocated that SMO is an important pairing with SEO because it helps users discover library content. As they noted, "Library SMO … enable[s] the growing number of users who engage in personalized searches, through their social media accounts, to be able to find library content by making it easier to link to the content, and by creating means for users to share the content" (2013, 105). Again, the novelty of this current article is the joint application of SEO and SMO for a coordinated effort and comprehensive understanding of libraries by social media networks and search engines.

Linked data and structured data practices for collections

Byrne and Goddard (2010) addressed the many potential benefits of libraries using linked data to describe their content, including disambiguation of information and understanding of content. Social media networks may also send signals to search engines, so optimizing library content for these environments may help search engines push relevant information to users. Scott (2015) described his library's use of White Hat SEO, which entails using structured data to have website information discoverable and properly understood by search engines. His approach extends to include the ways in which structured data can clarify how content should display in social media networks. Smith-Yoshimura (2016) noted the experimental nature of and lack of conclusive results from applying linked data to library projects. This state of libraries experimenting with linked data inspired us to see how application to paid resources would influence traffic to this content to add to the professional conversation about linked data use. Guha, Brickley, and Macbeth (2015) suggested that the shared vocabularies offered by Schema.org and linked data offers a structure well-understood by search engines beyond the traditional schema used by libraries. We adapted these ideas in the application of linked data and Schema.org markup to our Open SESMO project.

Methodology

As mentioned above, Open SESMO methods are based on several components: structured semantic markup, SEO and SMO, and employing linked data techniques. The first component (creating structured semantic markup and machine-readable data) involves two related techniques. The first technique is adding semantic HTML tags into your pages that help define structure and outline of the page for humans and machines. For example, one might add a `<main>` tag that wraps and identifies the main content present in the page or a `<header>` tag that wraps and identifies the header content of the page. The second technique is extending this tagging using additional markup attributes in the HTML that actually classify, define, or point to additional descriptions about the content contained in that HTML tag. For example, our `<main>` tag might get extended to use an external vocabulary type or property, and could now look like this: `<main type of="Product" property="mainEntity">`. This second technique applies RDFa (Resource Description Framework in Attributes) which "provides a set of markup attributes to augment the visual information on the Web with machine-readable hints."[1] It is here where the second component of Open SESMO (SEO and SMO) come into play. In addition to common SEO and other optimization techniques that sure a website can be crawled, indexed, and displayed in search engines or social media applications, an emerging practice has been to use a structured data markup schema that "helps search engines understand the information on Web pages and provide richer search results."[2] Schema.org, a collaborative effort between Google, Microsoft, Yahoo!, and Yandex, is one of these

structured data markup schemas that provides a web vocabulary for tagging web pages. Referring back to the earlier example, `<main typeof="Product" property="mainEntity">`, we are using the standard of RDFa to implement Schema.org types and properties into our HTML tagging and markup practices. These types and properties are understood by search engines and help them classify and index web pages with a more refined understanding of the content that can help with machine interpretations on the "aboutness" of the page and enable richer search results. The final component of Open SESMO follows the practice and concepts behind linked data, a set of best practices for publishing and connecting structured data on the web. In our case, we used a URL to link to a machine-readable description or extension to introduce additional descriptive vocabularies into our web pages. If we pick up again with our earlier `<main>` example, we can add a machine-readable definition of the content contained within the tag. For example, if the content inside of the `<main>` tag is a description of a database, and we want the machine-readable concept of database to be linked to the tag (and the content within), we would add another RDFa attribute named resource and use the URL to link to that external definition. The revised `<main>` tag would now look like this:

```
<main typeof="Product" property="mainEntity"
resource="http://dbpedia.org/resource/Database">.
```

This technique connects our pages to the linked open data on the web and sets a meaning that can be interpreted by machines and intelligent agents to discover the concepts and ideas in our web pages.

With a background discussion of the components of Open SESMO in place, we will walk through our application of each component here to outline the steps one can follow to apply the Open SESMO methodology. As we mentioned earlier, we had implemented similar optimization and HTML markup enhancement work with our digital collections, and we posited that this work could be applied to other types of web content within the library. In this case, the structured data HTML markup was applied using RDFa and generated from an SQL database that populated a web page template, making the practice scalable once we had settled on the necessary Schema.org types and properties. The four steps in our method are listed below.

(1) Apply established and reputed SEO techniques to build a site architecture that enables web crawling, site indexing, and SMO.
(2) Codify the relationship between the pages using structured semantic markup and machine-readable data.
(3) Designate each item (profile) page for a library database or resource as a Product entity using the Product type from Schema.org.
(4) Encode subject and topic vocabularies into all relevant pages for humans and machines using the External Enumerations, an extension that allows for external vocabularies to be encoded and referenced in an HTML page.

A first step in the method is building a site architecture, a list of pages for indexing, and additional markup in accordance with search engine and SMO techniques. To this end, the components of SEO that we followed were threefold: a site architecture built with indexing in mind, setting preferred URLs to be surfaced in SERPs, and building a site inventory for search engines using the sitemap.org protocol. It should be noted that our optimizations here follow the ethical SEO or "White Hat" SEO model and implement only well-established and supported SEO techniques that are honored by commercial search engines. In practice, this work involved establishing URL patterns, and setting canonical URLs for each page along with an XML sitemap for indexing. The website architecture was built to allow bots and web crawlers access to the broadest range of pages. Specifically, we set up a title list page that allows humans and machines access to each of the database profile pages. We also used a robots.txt file that directed bots and web crawlers into only the web directories necessary to build a functioning search index for the site. In establishing URL patterns and preferred URLs, we turned to the `<link> rel=canonical` markup that would tell search engines which URL to list in SERPs. The `<link> rel=canonical` markup appears in the `<head>` section on every page. You can see it in practice below.

```
<link property="significantLink" rel="canonical"
href="https://www.lib.montana.edu/resources/about/116"/>.
```

The href attribute supplies a value to bots and crawlers indicating that this URL is the one to use when surfacing an item in a SERP. The final practice of SEO that we followed was in establishing an index of URLs for search engines through a dynamically-generated sitemap. This XML file follows a protocol set by commercial search engines for listing the pages that you would like to see in their index of your site. Once we were satisfied with the sitemap.xml listing, we submitted the file through the Google Search Console to notify the search engine that our site was ready for indexing. In applying SMO, we followed the developer site guidelines for Twitter and Facebook to add Twitter Cards and Facebook Open Graph markup to each of the database item pages. For the Twitter Card tagging, we selected the Large Image Summary Card type, added the meta tags to each page, and then used the Twitter Card Validator to verify that we had made our changes correctly. For Facebook, we also added meta tags to each page and then used the Open Graph Debugger to validate that the Facebook Crawler correctly interpreted our tags. Examples of this social media markup best practices would look like this in a website:

```
<meta property="og:title" content="Mergent Online"/>
<meta property="og:description" content="Mergent Online
provides access to company profiles and financial information
with full text business profiles, quarterly and annual finan-
cials, EDGAR filings, and recent press releases and news wire
reports."/>
```

```
<meta property="og:image"
content="https://www.lib.montana.edu/resources/meta/
img/screenshots/116.jpg"/>
<meta property="og:url"
content="https://www.lib.montana.edu/resources/about/
116"/>
<meta property="og:type" content="website"/>
<meta name="twitter:creator" property="og:site_name"
content="@msulibrary"/>
<meta name="twitter:card" content="summary_large_image"/>
<meta name="twitter:site"
content="https://www.lib.montana.edu"/>
```

The second step in the method involves using structured semantic markup and machine-readable data practices to codify the relationships between a website's pages. Two of the entity types in Schema.org are Website and WebPage, which can have properties expressing the connections and relationships between them. The specific WebPage types that we used to classify our pages were

- CollectionPage: The landing and homepage;
- AboutPage: The topics page listing and the about page describing the collection; and
- ItemPage: The actual database profile page listing its qualities and the link to the database itself.

The type markup appears early in the `<body>` tag of the pages in a set of HTML attributes. Here is an example of the markup designating an ItemPage.

```
<body id="item" typeof="ItemPage"
resource="https://www.lib.montana.edu/resources/about/
116#item">
<meta property="name"content="Mergent Online"/>
<meta property="thumbnailUrl"
content="https://www.lib.montana.edu/resources/meta/img/
screenshots/thumb-116.jpg"/>
<meta property="isPartOf" typeof="CollectionPage"
resource="https://www.lib.montana.edu/resources/"
content="https://www.lib.montana.edu/resources/"/>
```

Note how the type is also given a resource address as one of the attributes. This URL value allows us to point machines at the content and reuse the data as a linked data resource. You can also see how the relationship between this ItemPage and the landing page are assigned in the `<meta property="isPartOf">`, which establishes this item page *as part of* a larger collection. The other ItemPages and AboutPages have similar markup assigning and linking the pages together as a whole collection.

The third step is designating each item (profile) page for a library database or resource as a Product entity using the Product type from Schema.org. This step is necessary for getting a particular kind of display into SERPs. Each ItemPage is given a mainEntity and identified as a type of Product. The markup for this designation is below.

```
<main id="product"typeof="Product"property="mainEntity
"resource="https://www.lib.montana.edu/resources/about/
116#product">
<h2 class="mainHeading"property="name">
Mergent Online</h2>
<img alt="Mergent Online screenshot" property="image"
src="https://www.lib.montana.edu/resources/meta/img/
screenshots/thumb-116.jpg" />
```

You can see the Product type and the mainEntity attributes making this assignment. This is essential work for describing the things on our pages as a type of product that can have particular properties like ratings, categories, creators, and so forth, and a particular layout in SERPs. The ratings and metadata visible in Open SESMO items on a typical SERP are possible because of this markup as seen in Figure 1 below.

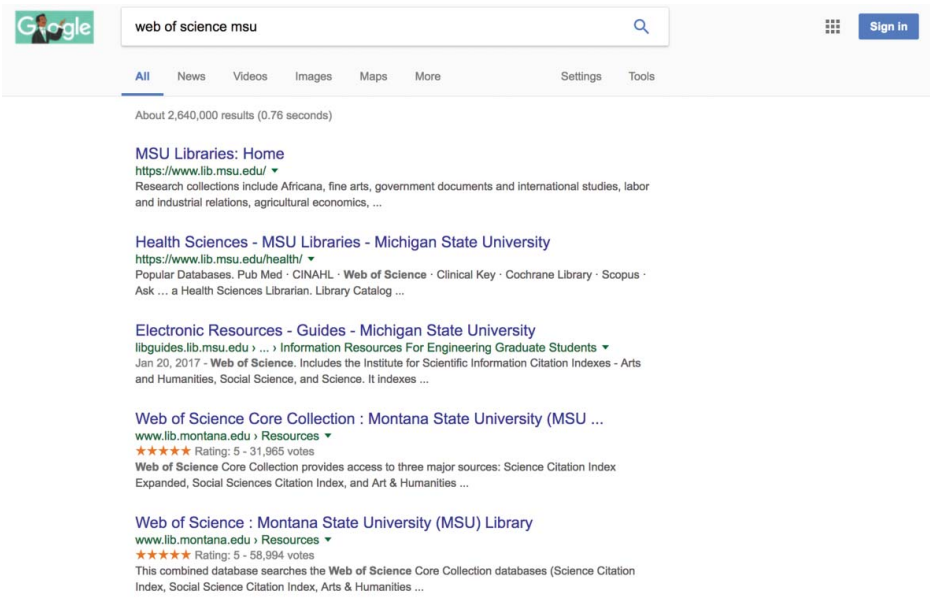

Figure 1. Screenshot showing Open Sesmo ItemPage in a SERP with ratings and enhanced visibility.

A fourth, final step in the method is setting topics and subjects for the pages by using linked open data vocabularies to encode subject and topic vocabularies into all relevant pages. In this linked data work, we turn to an emerging technique for assigning external vocabularies to our pages called "external enumerations." External

enumerations are a means to link to and make direct use of externally-maintained vocabularies and datasets using a resource attribute or <link> tag in the HTML of a web page. Sample markup for topics on one of our ItemPages is below.

```
<dl>
<dt><strong>Linked Data Topics:</strong></dt>
<dd property="additionalType"
resource="http://dbpedia.org/resource/Accounting"><a
property="additionalType"
href="https://en.wikipedia.org/wiki/Accounting"><span
property="category">Accounting</span></a></dd>
<dd property="additionalType"
resource="http://dbpedia.org/resource/Business"><a prop-
erty="additionalType"
href="https://en.wikipedia.org/wiki/Business"><span
property="category">Business</span></a></dd>
</dl>
```

In this case, the markup is showing HTML with external enumerations markup pointing to a DBpedia linked data resource. We chose DBpedia, a crowd-sourced community effort to extract structured information from Wikipedia, as our subject and topic vocabulary. We preferred its centrality in the linked open data cloud and the way in which search engines use the structured data from DBpedia to enhance SERPs and knowledge panels which you can see in Figure 2 below.

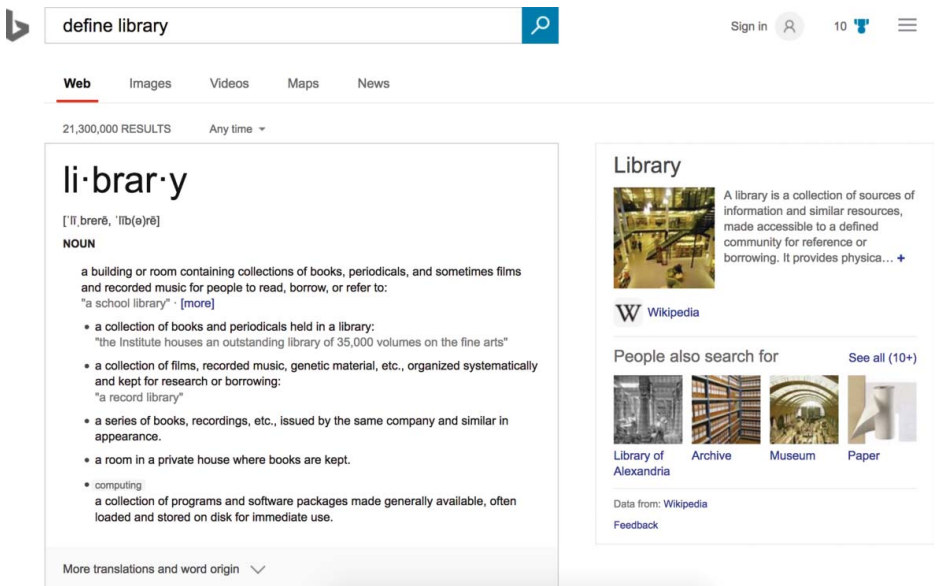

Figure 2. Screenshot showing BING Knowledge panel for "Library" using Wikipedia and DBpedia data as a source.

To find a set of vocabulary terms, we searched potential topics and subjects in Wikipedia, and added the identifier URLs to our SQL database for use in our web pages. We view this practice as a form of "web-scale cataloging" where description occurs and has an impact in classifying and enabling discovery of library resources beyond our local tools like the Integrated Library System (ILS).

Implementation of changes and results

In order to understand the impact of our changes, we began with a basic analysis of our website analytics for our library resources and indexes pages. These pages were consistently used and commonly among the top ten pages within our website. A snapshot from Google Analytics, visible in Figure 3 below, shows the common metrics focusing on "Acquisition" and "Behavior" that we used as a baseline.

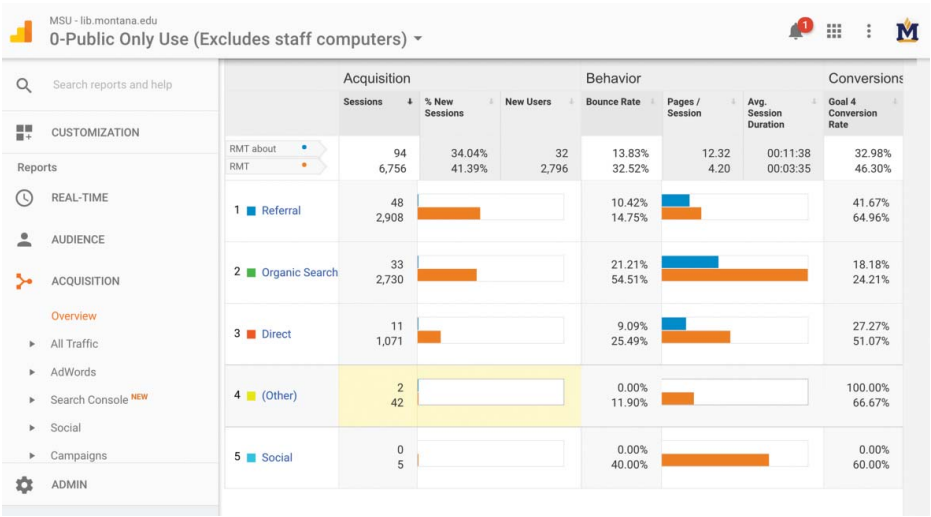

Figure 3. Screenshot from Google Analytics showing common traffic metrics for the whole resources subsite and the resources item pages prior to optimization 1 July to 21 August 2015.

Additionally, we looked at Facebook Analytics and Twitter Analytics to gauge interaction with social media posts which had been optimized for those networks. We also used the Google Search Console to monitor the search terms, search position, clickthrough rate, and common queries associated with the pages.

We implemented the changes to our website over time, with the last of these being made in time for the start of the fall semester. The results of the changes showed up gradually since search engines take time to update their indices. It was obvious once the pages were indexed as we saw a broad impact on traffic and referrals to our databases and resources pages. One of the first metrics that we saw was a significant jump in the impressions of our pages within commercial search engines (based on analytics in the Google Search Console). An impression occurs when a link appears in a search result for a user and denotes an increased visibility in

organic search results. To get a sense of the pre- and post-results, we recorded snapshots that straddled the implementation date of 22 August 2015. From these initial snapshots taken for 16 August 2015 to 12 September 2015 versus 13 September 2015 to 10 October 2015, we registered a growth in total page impressions from 14,273 to 55,292 as seen in Figure 4 below.

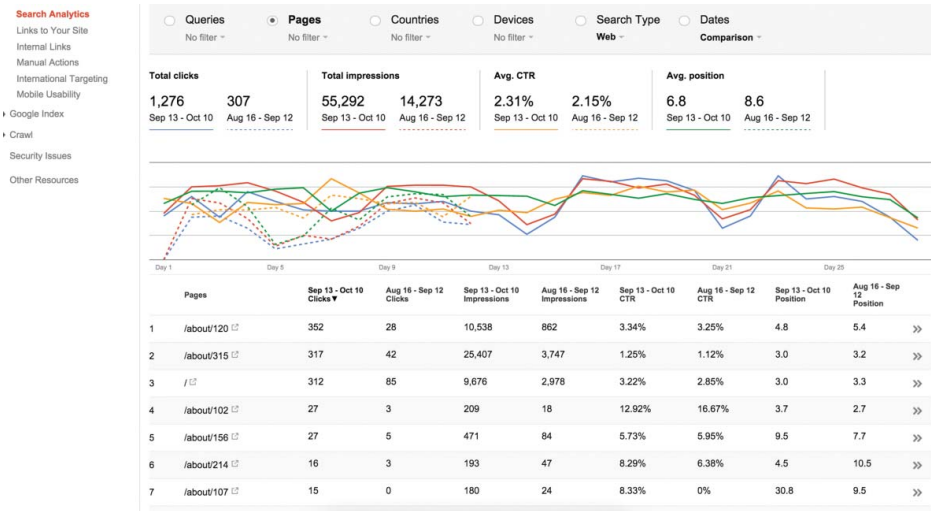

Figure 4. Screenshot from Google Search Console showing Initial Snapshot for Impressions, Clickthrough Rates (CTR), and Search Position.

This impression growth shows that the number of times our library resource items were being seen in SERPs increased at almost a 400 percent rate. Related search engine metrics such as "Clickthrough Rates" (CTR) and "Search Position" showed growth as well. In the CTR and Search Position results, we see our items appearing higher up in search results rankings and this higher placement of our items potentially leading to new levels of engagement where the user is actually clicking through from the search result page to our library resources landing pages. It should be noted that higher rankings were not a direct goal as placement and pursuit of a #1 ranking can be subject to some of the more obtuse search algorithms. However, research has shown that items that appear within the first page of search results "garnered 92 percent of all traffic from the average search, with traffic dropping off by 95 percent for Page 2," so there is reason to monitor the impact of a higher ranking.[3]

In addition to impressions, we could start to quantify organic search growth by using Google Analytics to survey changes over three distinct fall semesters (i.e., 2014, 2015, and 2016) for our about/item pages.[4] These about/item pages are a small subset of pages comprised of the optimized landing pages where we placed the new markup. As such, the numbers mentioned here are looking at around 260 pages that we put into the index for commercial search engines and do not represent the complete MSU Library Resources website. Within the Google

Analytics – Acquisition, Source/Medium[5] module, we watched our Google/Organic Source/Medium analytics move from 104 sessions[6] in 2014 to 5,698 sessions in 2015, and settle at 5,561 sessions in 2016.

Year	Organic search sessions
2014	104
2015	5,698
2016	5,561

This change in sessions confirms the rising engagement levels with our pages as a result of search traffic. Individual user growth also showed steady progress, moving from 104 users in 2014 to 4,190 users in 2015, and settling at 3,175 users in 2016.

Year	Number of individual users
2014	104
2015	4,190
2016	3,175

This user metric shows steady growth in the number of visitors referred by search engines to our about/item pages. We can also start to see some changing user behavior patterns. Along these lines, we noted a change in our "Exit Rates" as the exit rate percentages shifted from 45 percent to 53.16 percent over the 2014 to 2016 time period.[7] These are healthy exit rates that indicate that the links into our subscription resources on the optimized landing (about/item) pages are leading users from the page to our actual library resources. We are also able to get a sense of the browsing activity of our users in the "Pages per Session" metrics.

Year	Pages per session
2014	26.59
2015	1.94
2016	3.60

The decrease from an unwieldy 26.59 pages/session in 2014, reset into 1.94 pages/session in 2015 and settling into a preferred 3.60 pages/session ratio indicates a healthy browse point in leading users into additional library resources, but not getting them lost (which is what the 26.59 pages/session suggests). Improvements in user browsing behavior are also apparent in the "Average Session Duration" metric.

Year	Average session duration (minutes:seconds)
2014	00:09:43
2015	00:01:37
2016	00:02:35

In this metric, we saw the time users spent browsing our pages move from an unhealthy 9 minutes and 43 seconds in 2014 to 1 minute and 37 seconds in 2015, and end at 2 minutes and 36 seconds in 2016. We note this as a healthy trajectory because the purpose of the site is to refer people to information about a subscription resource, let them look at a few related resources, and then elect into loading the subscription resource. The 9 minutes and 43 seconds we were seeing in 2014 was not accomplishing these site goals and suggests people were browsing at the expense of making a decision. Beyond the user experience metrics, we saw some of the most telling and dramatic changes in our "Pageviews."[8]

Year	Pageviews
2014	2,765
2015	11,018
2016	19,878

For the about/item pages, we witnessed our total pageviews move from 2,765 in 2014 to 11,018 in 2015, and finish at 19,878 in 2016. In analyzing all of these metrics, it became apparent that we were seeing impactful growth and results from the Open SESMO techniques and optimizations.

For our social media posts, we generally see little to no interaction with posts about databases when the posts are text-only and not optimized for these networks. With our SMO, posts about library paid databases pull in information about the databases including a screenshot of the database homepage, a brief description of the resource, and a link to the library's Facebook or Twitter page. Analytics such as Twitter Analytics offer metrics such as impressions (the number of time people saw your post), "clickthroughs" to the link shared, likes, retweets, and so on. Figure 5 shows an optimized database shared on Twitter, and figure 6 reveals interactions with this Tweet.

Discussion

In embarking on the Open SESMO project, we sought to make our databases visible to our users and to be clearly understood by search engines as being connected with our institution. Our efforts and the efforts of any library applying these same techniques will produce different results depending on a number of factors, including geographical location of the search being conducted, the web browser used (e.g., Chrome, Safari), user cookies and search history, terms searched, search engine used (e.g., Google, Bing), and terms on your website. Our goal was to clearly surface our resources to our users and to disambiguate our database from other more generic descriptions of these resources. We also note that we followed ethical SEO practices that are clearly supported and encouraged by search engines. These best practices (e.g., structured data, crawling directives, semantic HTML, Open Graph markup) are

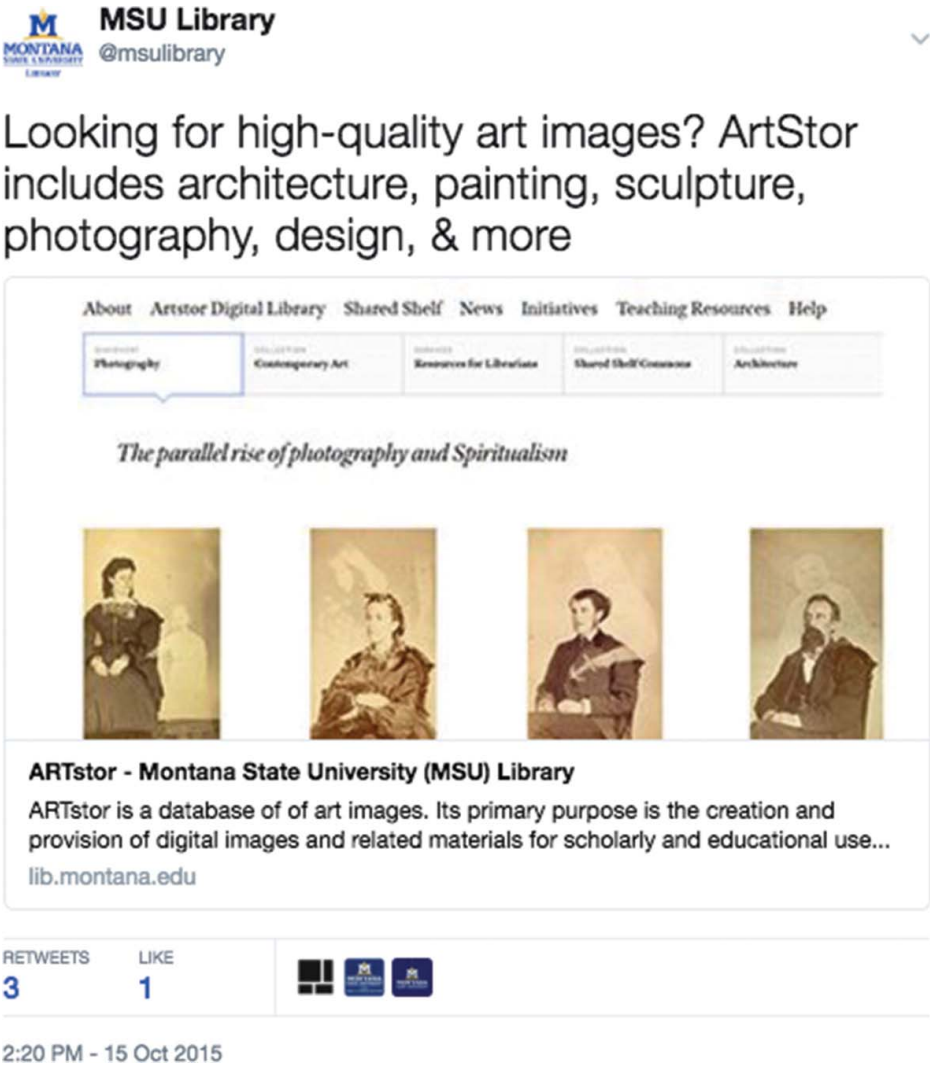

Figure 5. Screenshot from Twitter showing Tweet of database optimized for social media sharing.

bedrock activities that have a longevity and history of support within commercial search engines and social media applications. We implemented and built Open SESMO around these foundational techniques to avoid obsolescence and ranking penalties that might result from attempts to "game the system."

Over the course of the project, we discovered a number of limitations and challenges. First, we note the variability of the searchable index within commercial search engines. We were able to watch the growth and changes to our indexed pages in applying the "site:" operator to various search engines and taking snapshots using Zotero to document the progress. We also used the Google Search Console to monitor crawl errors and the health of our searchable index. Depending on the search engine, we noted a larger searchable index and different ranking

Impressions	1,773
Total engagements	17
Link clicks	6
Detail expands	6
Retweets	3
Likes	1
Profile clicks	1

Figure 6. Screenshot from Twitter Analytics showing interactions with Tweet of database optimized for social media sharing.

rules for our Open SESMO content. You can see some of these differences in the screenshots in Figures 7 and 8 below, which compare the same "site:" operator search in Google to Bing.

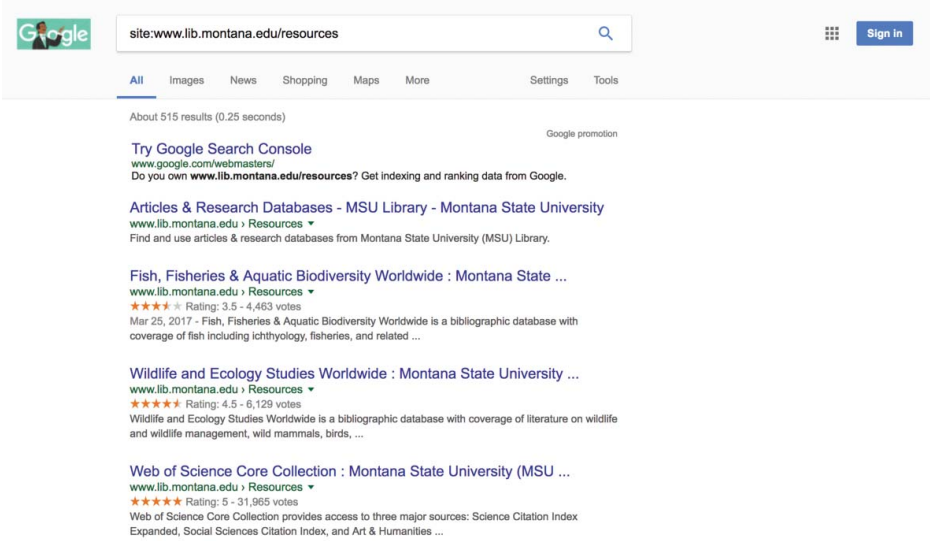

Figure 7. Site operator query for www.lib.montana.edu/resources in Google.

It was clear that each search engine used different markers for relevance and interest for a particular user. Additionally, search engines adjust ranking algorithms and rules for inclusion in the index to keep a rewarding user experience in place. At times, this can feel like working within a game where the rules keep changing. We encourage practitioners who are looking to follow our methods to allow time and learning for keeping up to date on SEO and SMO practices as methods are often in flux. However, we did find that even within this dynamic environment, the core SEO and SMO strategies and structured, linked data markup in our Open SESMO methods consistently performed well and provide a baseline of

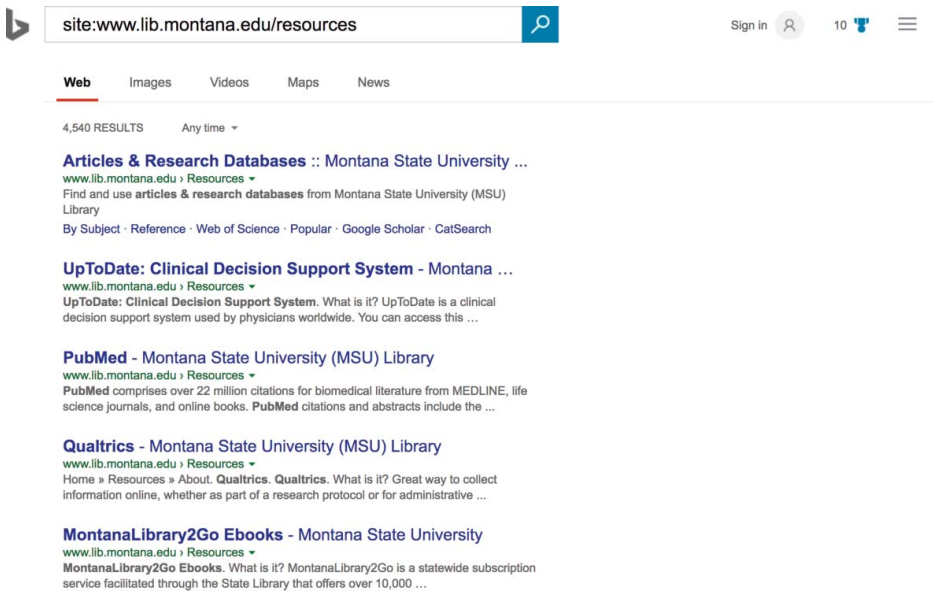

Figure 8. Site operator query for www.lib.montana.edu/resources in Bing.

optimization techniques for libraries moving forward. Second, we note the need to expand Open SESMO techniques into content management systems like LibGuides to enable libraries to have their content understood better by search engines and, therefore, more discoverable by their users. Vendors could allow customization of their sites, specifically the markup, attributes, and optimizations we feature in our Open SESMO method, to be applied to their products for each library customer. In addition, vendors could provide libraries with better statistics about how our users find our resources using similar data to what is provided by Google Analytics with information such as referrer, exit rates, bounce rates, and so on. Third, we note the intense and deliberate testing/experimentation around structured data markup, SEO, and SMO has a high learning curve, needs institutional investment, and requires frequent monitoring of changes to search and social platforms. In our library, we have several initiatives concerned with Semantic Web work and optimizing for being found in external platforms like social media and search engines. We have integrated these initiatives into several research threads and have used the results we are seeing to make broader cases for library support to library and university administration. Approaching these structured data activities with clear goals in mind will help give your local initiative some momentum and create a broader impact for your library.

Keep in mind that the database pages in the Open SESMO project are targeted for understanding by search engines and social media networks. These are pages set up to feed the machine: the bots and intelligent software agents looking to index and understand our site. We do also rely on our Springshare LibGuides for library instruction, research consultations, and additional browsing contexts for

subscription databases and library guides. The LibGuide listings route through our permanent Open SESMO URLs for databases so we can capture as much database traffic as possible through Google Analytics. It would be ideal for these sites to be combined into one to reduce redundancy in maintenance, but the LibGuides platform does not allow for the level of customization around markup and optimization present in our Open SESMO project. User experience testing for our item-level database pages would better inform what people understand about our website once they go to resources from search engine results and could further shape how we present these resources. As for where continued work on the Open SESMO project will focus, we have begun looking at several themes: investigation into emerging formats of structured data, research into what makes a searchable index, and local Search Engine and SMO.

As the final part of this discussion, we outline a series of steps for thinking about Open SESMO in your local library web pages. It is not quite a roadmap to implementation, but it does provide a sense of the thinking that is necessary to set up the method for a local effort.

(1) Think about how your content on your pages might be identified and grouped using semantic HTML tags.
(2) Study and apply Schema.org vocabulary for types and properties relevant for defining your pages. Use RDFa or another means to embed the structured data on your pages.
(3) Browse and search Wikipedia for common terms you might use to describe your pages. Link your pages using RDFa resource attributes to machine-readable linked data descriptions of your content using equivalent links in DBpedia.[9]
(4) Set up a sitemap file that lists the pages (URLs) you would like to see crawled along with their priority ranking relative to the other URLs listed in the sitemap.
(5) Survey local user groups to see where their social media usage is centered. Apply relevant social media markup based on your findings (see Rossmann and Young 2016).
(6) Connect your pages to an analytics software package, and monitor analytics to gauge the results of the change(s) to your library's website.

Conclusion

In our results and case study of Open SESMO, we show a demonstrable benefit to bringing library databases and resources into SERPs. In configuring the SERP as an interface for the library, we saw increased visitation and use of our library databases and resources. Beyond our findings that optimization and structured, linked data has an impact on discovery in search engines, we see that our method introduces a new teaching moment for libraries: meeting our users where they are, in the commercial search interface and on social media networks. We were also struck by the

fact that these optimization techniques came from our digital library collection practices and further demonstrate the need for our library specialties to work in tandem. We see what we are calling "Web-scale Cataloging" (i.e., the central techniques of Open SESMO: semantic markup, structured data, and linked data vocabularies) as part of an emerging teaching moment and a new skillset for librarians.

Notes

1. The World Wide Web Consortium (W3C) defines and maintains the RDFa standard at https://www.w3.org/TR/xhtml-rdfa-primer/.
2. A complete FAQ for Schema.org is available at https://schema.org/docs/faq.html.
3. "No. 1 Position in Google Gets 33% of Search Traffic" [Study]. https://searchenginewatch.com/sew/study/2276184/no-1-position-in-google-gets-33-of-search-traffic-study.
4. Organic Search Growth is the measurement of referred traffic from search engines.
5. The Source is the place users are before seeing your content, like a search engine or another website. The Medium describes how users arrived at your content.
6. A Session is the period time a user is actively engaged with your website, app, etc.
7. Exit Rate as a term used in website traffic analysis (sometimes confused with bounce rate) is the percentage of visitors to a site who actively click away to a different site from a specific page.
8. A Pageview is recorded every time a page is viewed.
9. A Wikipedia page can be turned into a DBpedia page using this convention: https://en.wikipedia.org/wiki/Database => https://dbpedia.org/resource/Database. The essential term for the mapping is the last bit of the Wikipedia URL that identifies the page.

Notes on contributors

Jason A. Clark (jaclark@montana.edu), is an Associate Professor and Head of Library Informatics and Computing at Montana State University, specializing in web development, metadata and data modeling, web services and APIs, search engine optimization, and interface design. You can find him online at http://www.jasonclark.info and as @jaclark on Twitter.

Doralyn Rossmann (doralyn@montana.edu), is an Associate Professor and Head of Collection Development at Montana State University. Her research interests include social media community building, ethics, and optimization; public budgeting; and organizational culture. She is @doralyn on Twitter and can be found at http://doralyn.org.

ORCID

Jason A. Clark http://orcid.org/0000-0002-3588-6257
Doralyn Rossmann http://orcid.org/0000-0002-6490-4223

References

Alford, Emily. 2009. Promoting and Marketing E-Resources. *The Serials Librarian* 57 (3):272–277. Accessed March 18, 2017. https://doi.org/10.1080/03615260902877084. doi:10.1080/03615260902877084.

Arlitsch, Kenning, Patrick O'Brien, and Brian Rossmann. 2013. Managing Search Engine Optimization: An Introduction for Library Administrators. *Journal of Library Administration* 53

(2–3):177–188. Accessed March 26, 2017. https://doi.org/10.1080/01930826.2013.853499. doi:10.1080/01930826.2013.853499.

Arlitsch, Kenning, Patrick O'Brien, Jason A. Clark, Scott W. H. Young, and Doralyn Rossmann. 2014. Demonstrating Library Value at Network Scale: Leveraging the Semantic Web with New Knowledge Work. *Journal of Library Administration* 54 (5):413–425. Accessed March 26, 2017. https://doi.org/10.1080/01930826.2014.946778. doi:10.1080/01930826.2014.946778.

Barghava, Rohit. 5 Rules of Social Media Optimization. Accessed March 18, 2017. http://www.rohitbhargava.com/2006/08/5_rules_of_soci.html.

Barghava, Rohit. The New Rules of Social Media Optimization. Accessed June 11, 2017. http://www.rohitbhargava.com/2010/08/the-5-new-rules-of-social-media-optimization-smo.html.

Belden, Dreanna. May 2007. Are you using Wikipedia to draw users to your online materials?" *Library 2.0*. Accessed March 18, 2017. http://www.library20.com/group/governmentdocuments/forum/topic/show?id=515108%3ATopic%3A29392.

Byrne, Gillian, and Lisa Goddard. 2010. The Strongest Link: Libraries and Linked Data. *D-Lib Magazine* 16 (11–12). Accessed March 28, 2017. https://doi.org/10.1045/november2010-byrne.

Dugan, Mary. 2011. 'Database of the Week': Successfully Promoting Business Databases to Faculty. *Journal of Business & Finance Librarianship* 16 (2):159–170. Accessed March 18, 2017. https://doi.org/10.1080/08963568.2011.556351. doi:10.1080/08963568.2011.556351.

Evgenia, Vassilakaki, and Emmanouel Garoufallou. 2014. The Impact of Facebook on Libraries and Librarians: A Review of the Literature. *Program* 48 (3):226–245. Accessed March 19, 2017. https://doi.org/10.1108/PROG-03-2013-0011. doi:10.1108/PROG-03-2013-0011.

Fry, Amy. 2014. The Biggest Winner: Using Statistics to Assess the Effectiveness of an E-Resources Promotional Campaign. *Journal of Electronic Resources Librarianship* 26 (1). Accessed March 19, 2017. https://doi.org/10.1080/1941126X.2014.877330. doi:10.1080/1941126X.2014.877330.

Gauder, Brad. 2011. Perceptions of Libraries, 2010: Context and Community. A Report to the OCLC Membership. OCLC Online Computer Library Center, Inc. Accessed March 7, 2017. http://files.eric.ed.gov/fulltext/ED532601.pdf.

Guha, R. V., Dan Brickley, and Steve Macbeth. 2015. Schema.org: Evolution of Structured Data on the Web. *Queue* 13(9). Accessed March 28, 2017. http://queue.acm.org/detail.cfm?id=2857276. doi:10.1145/2857274.2857276.

Gustafson, Julia Chance, Zachary Sharrow, and Gwen Short. 2017. Library Marketing on a Small Liberal Arts Campus: Assessing Communication Preferences. *Journal of Library Administration* 57 (4):420–435. Accessed June 11, 2017. https://doi.org/10.1080/01930826.2017.1300459. doi:10.1080/01930826.2017.1300459.

Hilliker, Robert J., Melanie Wacker, and Amy L. Nurnberger. 2013. Improving Discovery of and Access to Digital Repository Contents Using Semantic Web Standards: Columbia University's Academic Commons. *Journal of Library Metadata* 13 (2–3):80–94. Accessed March 26, 2017. https://doi.org/10.1080/19386389.2013.826036.

Kwanya, Tom, Christine Stillwell, and Peter G. Underwood. 2014. *Library 3.0: Intelligent Libraries and Apomediation*. Oxford, UK: Elsevier.

Monie, Christine, and Jessica Clark. 2013. Promoting the Library through an Electronic Table of Contents (e-TOC) E-mail Service: The Wollongong Experience. *Journal of Hospital Librarianship* 13 (1):32–41. Accessed March 18, 2017. https://doi.org/10.1080/15323269.2013.743206. doi:10.1080/15323269.2013.743206.

Noh, Younghee. 2015. Imagining Library 4.0: Creating a Model for Future Libraries. *The Journal of Academic Librarianship* 41 (6):786–797. Accessed August 1, 2017. https://doi.org/10.1016/j.acalib.2015.08.020. doi:10.1016/j.acalib.2015.08.020.

Onaifo, Daniel, and Diane Rasmussen. 2013. Increasing Libraries Content Findability on the Web with Search Engine Optimization. *Library Hi Tech* 31 (1). Accessed August 1, 2017. https://doi.org/10.1108/07378831311303958 doi:10.1108/07378831311303958.

Rossmann, Doralyn, and Scott W. H. Young. 2016. Social Media Optimization: Principles for Building and Engaging Community. *Library Technology Report* 52 (8). Accessed March 18, 2017. https://doi.org/10.5860/ltr.52n8.

Schonfeld, Roger. 2015. Meeting Researchers Where They Start: Streamlining Access to s\Scholarly Resources. Ithaka S R. Accessed March 7, 2017. http://sr.ithaka.org/blog-individual/meetingresearchers-where-they-start-streamlining-access-scholarlyresources.

Scott, Dan. 2015. White Hat Search Engine Optimization: Structured Web Data for Libraries. *Partnership: The Canadian Journal of Library and Information Practice and Research* 10 (1). Accessed June 11, 2017. https://doi.org/10.21083/partnership.v10i1.3328.

Smith-Yoshimura, Karen. 2016. Analysis of International Linked Data Survey for Implementers. *D-Lib Magazine* 22 (7–8 July/August). Accessed March 28, 2017. https://doi.org/10.1045/july2016-smith-yoshimura.

Solodovnik, Iryna. 2011. Metadata Issues in Digital Libraries: Key Concepts and Perspectives. *Italian Journal of Library, Archives, and Information Science* 2 (2). https://www.jlis.it/article/view/4663.