



School of Law
UNIVERSITY OF GEORGIA

Prepare.
Connect.
Lead.

Digital Commons @ University of
Georgia School of Law

Articles, Chapters and Online Publications

Alexander Campbell King Law Library

11-25-2019

Here Come the Bots: Six Tips When Designing Your IR's Metadata for Improved Discoverability

Rachel S. Evans

University of Georgia School of Law, rsevans@uga.edu

Repository Citation

Evans, Rachel S., "Here Come the Bots: Six Tips When Designing Your IR's Metadata for Improved Discoverability" (2019). *Articles, Chapters and Online Publications*. 56.

https://digitalcommons.law.uga.edu/law_lib_artchop/56

This Article is brought to you for free and open access by the Alexander Campbell King Law Library at Digital Commons @ University of Georgia School of Law. It has been accepted for inclusion in Articles, Chapters and Online Publications by an authorized administrator of Digital Commons @ University of Georgia School of Law. [Please share how you have benefited from this access](#) For more information, please contact tstriepe@uga.edu.

TSSL TechScans

A Blog to share the latest trends and technology tools for Technical Services Law Librarians

MONDAY, NOVEMBER 25, 2019

Here Come the Bots: Six Tips When Designing Your IR's Metadata for Improved Discoverability



Last week I attended a webinar about "the science of discoverability". Although it was aimed at librarians working with institutional repository (IR) content, it was an excellent reminder that the many best practices I followed as a web developer for our law school's Drupal site were applicable not only with repositories but also with LibGuides (and any other pages we wanted Google to find). Here are six tips to deploy when designing metadata for the bots and increasing your site's discovery:

1. **Title Fields Are Important!** In fact they are perhaps the most important field of any object or event metadata in your repository. Working as a web developer this was something we struggled with when other users would create webpages. The title did not always match or identify the content. Later on they inevitable call or email to ask why it isn't showing in Google's search results when they put in keywords that they assume will retrieve their exact webpage. Almost always the keywords they wanted Google to identify and pull their exact page from were not in their page title field or URL. The same rings true for IR content. No matter how many other fields have the data or keywords, if the title doesn't it probably isn't good enough to be retrieved by Google (unless you have big bucks of course - then you can use Adwords to pay your way to the top of that results list as a sponsored items..but I doubt any of us have that kind of money for SEO, hah!).

Subscribe via RSS feed



[Subscribe in a reader](#)

Subscribe via email

Enter your email address:

Delivered by [FeedBurner](#)

TSSL column editor

Travis Spence

Contributors

Carol Collins

Rachel Evans

Annie Mellott

Lauren Seney

Keelan Weber

Travis Spence

Former Contributors

Andrea Rabbia

Ashley Moye

Caitlyn Lam

Chris Tarr

Corinne Jacox

Dan Blackaby

Elizabeth Geesey Holmes

Ellen McGrath

Elyssa Gould

Emily Nimsakont

Ismael Gullon

Jason LeMay

Jean Pajerek

Marlene Bubrick

2. **HUMAN-Readable Is Better.** This is not your library catalog. Your ILS is a (mostly) closed-off system. It was engineered by ONLY librarians who have strict cataloging rules passed down over decades of meticulous rules with a field for every single possible bit of data. IR's are not an ILS. In the same way Google is not your OPAC. They do not and will never function the same way. Sure, you can use some of the same operators, and you may even form similar strings in each of the search bars. The difference is that Google's algorithm is not a 100% known entity. Most of Google's users are performing natural language searches. Your I.T. or metadata librarian's cannot get into Google's back-end and tell it what you want, what fields to provides searches for, how to weight results and how to display your results list. Google's algorithm not only likes but wants HUMAN-readable, NOT machine-readable. Craft the content in your fields for any given item, event, or landing page with this in mind. The key here is not to overdo it!

3. **Don't Use Too Many Keywords.** This relates to the last sentence of the last tip - don't overdo it. In addition to not getting overly wordy or technical in your fields, the field to especially watch out for is keywords. In Digital Commons there is a nice keyword field. When I first started adding content to our repository I no doubt went overboard with more keywords than I should have. Although too few could hinder discoverability, if the keywords are on point and you have two to four of them that are appropriate you will hit a sweet spot with Google's crawl. But beware of using too many. Google and other search engines will actually ping or potentially ignore your content (and in some cases as the webinar warned your entire site) for using too many keywords. Excessive metadata makes it assume this content isn't valid. So just be careful here. This doesn't mean you should *never* use more than four keywords. There may be occasions when less just won't cut it. Perhaps that one article or conference you just loaded is particularly interdisciplinary and really needs more terms. Keeping the majority of your content with three keywords or less will get search engines to take you more seriously and those few instances where you decided to use more keywords won't throw up red flags like twelve keywords for every single items in your repository would.

4. **Frequency, Consistency & Longevity.** I can't count how often I was asked as a web developer when and how often Google would crawl our site. This is a mystery to most everyone, and while you can request through some of Google's Webmaster Tools for a re-crawl there is no guarantee the speed at which that will happen. One thing is for sure, you will be re-crawled more often the more frequently and consistently you update any site, no matter what site it is. Long periods of no activity may result in flagging you as a dead site so regular adding or refreshing of content is the key here. Another related factor is longevity. This is simply the idea that the longer a site exists the more time it has had to be crawled, to appear in search results, and as a result to increase site traffic. Then the cycle returns to the beginning since the more site visits you receive from organic Google searches the more your site should rise in the results list as your site and its content becomes more closely associated with a variety of searches over time. Obviously a brand new site will take time to get there, but after many repeats of this cycle with the help of

Patricia Turpening

Rachel Purcell

Yumin Jiang

Links

[AALL web site](#)

[OBS-SIS web site](#)

[Technical Services Law Librarian](#)

[TS-SIS web site](#)

Topics

[aall annual meeting](#) (1)

[acquisitions](#) (59)

[CALI](#) (1)

[CALIcon](#) (1)

[cataloging](#) (320)

[conferences](#) (2)

[digital images](#) (2)

[discoverability](#) (1)

[getting to know you](#) (36)

[government documents](#) (25)

[information technology](#) (262)

[local systems](#) (66)

[management](#) (77)

[metadata](#) (38)

[new members](#) (4)

[preservation](#) (74)

[processing](#) (17)

[professional development](#) (2)

[quick question](#) (4)

[repository](#) (5)

[SEO](#) (1)

[serials](#) (39)

[web archiving](#) (1)

Blog Archive

▼ 2019 (23)

▼ November (4)

[Here Come the Bots: Six Tips
When Designing Your I...](#)

[Getting to Know TS Librarians:
Joy Humphrey](#)

[Quick Question: What
conferences should I attend?](#)

[GLA Conference Review:
Workshop on Digitization fo...](#)

► October (1)

your frequent and consistent care and feeding this will happen naturally.

5. **Bots Like Quick Load Times.** So since we don't really know when Google or other search engine bots will pay us a visit, how can we make sure that when they do they are finding us at our best? Load times are one big indicator. I know, I know... but there are SO many cool and flashy things we could embed into our content, right? Is that snazzy High-Res image of the latest guest lecturer too much for Google? What about our Issuu flipbooks of scanned symposia programs, or the YouTube video of the three hour panel? Each bit of multimedia needs a different approach here. If your IR system has native streaming this will help cut down on additionally embedded load times. If not, you may need to choose what is more important - the load time or the media keeping your traffic on your site. If traffic isn't a major factor, load times will increase by hyperlinking to the media instead of placing it on the page itself. The same could be true for embedded flip-book style PDFs. For images, as long as you use best practices for the proper resolution on the web you should not have to choose between a crisp, quality image and fast load times. Use the right format for image and other media files (choose MP3's for online streaming instead of WAVs or AIFFs). If you want or need to offer the highest quality original files to site visitors, hyperlink to that file's location instead of providing at their point of entry. This will keep load times up and still give visitors the option of access and retrieval. In the end, the faster your content loads, the more quickly it can be indexed. Bots are impatient - they are bots! Make them wait too long and they just keep moving.

6. **Site Maps Are Critical, Especially for "Dead" Collections.** So your content is now in tip-top shape! It has excellent human-readable title fields and abstracts. It has good keywords, but not *too* many of them. You've even managed to build a beautiful page of content enhanced with multi-media, but you've been careful to follow best practices for these files and your load time is great. Now there is just one problem - this collection is an archive! It just so happens as a librarian you have created a collection of items that will never grow again because it is historical. How can you possibly be frequent and consistent with this set of data? Will Google eventually forget about you (even if the collection exists over a long period of time) because there is nothing to update? No! Not necessarily - this is where your site's skeleton, the trusty site map, comes into play. Depending on the system you are using a site map may be generated for you as you create new content. It never hurts to revisit this though. Particularly for sites that have been around over a long period of time, the site map (generated for you or created by someone else) may be pulling titles and other structural and organizational information that is either no longer accurate or appropriate, or perhaps it is just not as good as it should be. Revisit your site map every so often as a regular maintenance task. It is essentially an outline of your site and all that it contains, and as such can indicate where a collection or series title is not descriptive enough, is too descriptive, or is just not human-readable. Think back to tip #1 and #2 for human-readable fields (especially titles). Page summaries can help here as well. When you conduct a Google search, if a result appeared but had no description at all for the page are you going to take your chances with clicking through to

- ▶ [August](#) (1)
- ▶ [June](#) (5)
- ▶ [May](#) (1)
- ▶ [April](#) (4)
- ▶ [March](#) (2)
- ▶ [February](#) (2)
- ▶ [January](#) (3)

- ▶ [2018](#) (31)
- ▶ [2017](#) (52)
- ▶ [2016](#) (38)
- ▶ [2015](#) (43)
- ▶ [2014](#) (42)
- ▶ [2013](#) (38)
- ▶ [2012](#) (48)
- ▶ [2011](#) (76)
- ▶ [2010](#) (80)
- ▶ [2009](#) (56)
- ▶ [2008](#) (150)
- ▶ [2007](#) (51)

that result, or are you more likely to choose the result that tells you what you will find there? Make titles, related page summaries for what it is about, and if possible even URL strings make sense and describe what you will find there. Adjust your sitemap and related descriptive data as needed, and monitor how your site (hopefully) rises in results over time, as well as how your traffic (hopefully) increases over time.

Have more tips to share with TechScans readers that were not touched on here? What has worked for improving your website or repository's metadata, and how do you optimize your content for search engines? Share with us in the comments!

Posted by Rachel Evans at [10:26 AM](#)



Labels: [discoverability](#), [metadata](#), [preservation](#), [repository](#), [SEO](#)

No comments:

[Post a Comment](#)

Links to this post

[Create a Link](#)

[Home](#)

[Older Post](#)

Subscribe to: [Post Comments \(Atom\)](#)