

Search Engine Visibility for Library Resources



Zoe Dickinson

zoe.dickinson@dal.ca



Outline

- Introduction: why search engine visibility matters
- Methodology
- How visible are Canadian public library resources right now?
- Reasons behind visibility (or lack thereof)
- Technical barriers and solutions
- Discussion



All Images News Videos Maps More Search tools

About 301,000,000 results (0.45 seconds)

✓ [The Illegal Book on Sale - Indigo.ca](#)

Ad chapters.indigo.ca/[The-Illegal-Book](#)

by Lawrence Hill Now 40% Off at Indigo.ca! Free Shipping Over \$25.

Ratings: Website 10/10 - Service 10/10 - Selection 10/10 - Product quality 10/10

📍 Halifax, NS - (902) 423-6438 - Open today · 9:00 AM – 6:00 PM

[New Harry Potter Book](#)

[New & Hot Books](#)

[Up to 40% Off Bestsellers](#)

▶ ✓ [The Illegal - Lawrence Hill - Hardcover](#)

www.harpercollins.ca/9781554683833/the-illegal

This is the new underground: a place where tens of thousands of people deemed to be "illegal" live below the radar of the police and government officials.

✓ ['The Illegal,' by Lawrence Hill - The New York Times](#)

www.nytimes.com/2016/01/17/books/.../the-illegal-by-lawrence-hill.html

Jan 15, 2016 - A marathoner enters a wealthy country illegally in Hill's novel.

✓ [Review: Lawrence Hill's The Illegal is a twisting, intricately ...](#)

www.theglobeandmail.com › Arts › Books & Media › Book Reviews

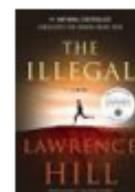
Sep 4, 2015 - In the afterword of his new novel, *The Illegal*, Lawrence Hill says it took him five years to write the book. I read the novel in less than a day.

✓ [The Illegal by Lawrence Hill — Reviews, Discussion ...](#)

www.goodreads.com › Cultural › Canada

Shop for the illegal on Google

Sponsored ⓘ



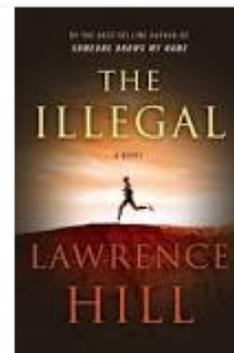
The Illegal

\$14.99 - Indigo Books & Music

The Illegal: A Novel



Book by Lawrence Hill



A fast-moving, epic work by internationally best-selling author Lawrence Hill about a man on the run. ... [Google Books](#)

Originally published: January 25, 2016

Author: [Lawrence Hill](#)

People also search for



What's missing from this picture?

- Many library resources do not appear in a typical search engine results page (SERP)
- As search engines become the ubiquitous “front page” of the internet, this poses a threat to libraries’ perceived relevance



A few relevant studies

- OCLC Perceptions of Libraries Report:
<http://oclc.org/reports/2010perceptions.en.html>
- Pew Internet Research Center's Internet and American Life Project: <http://www.pewinternet.org/2012/03/09/main-findings-11/>



What do I mean by visibility?

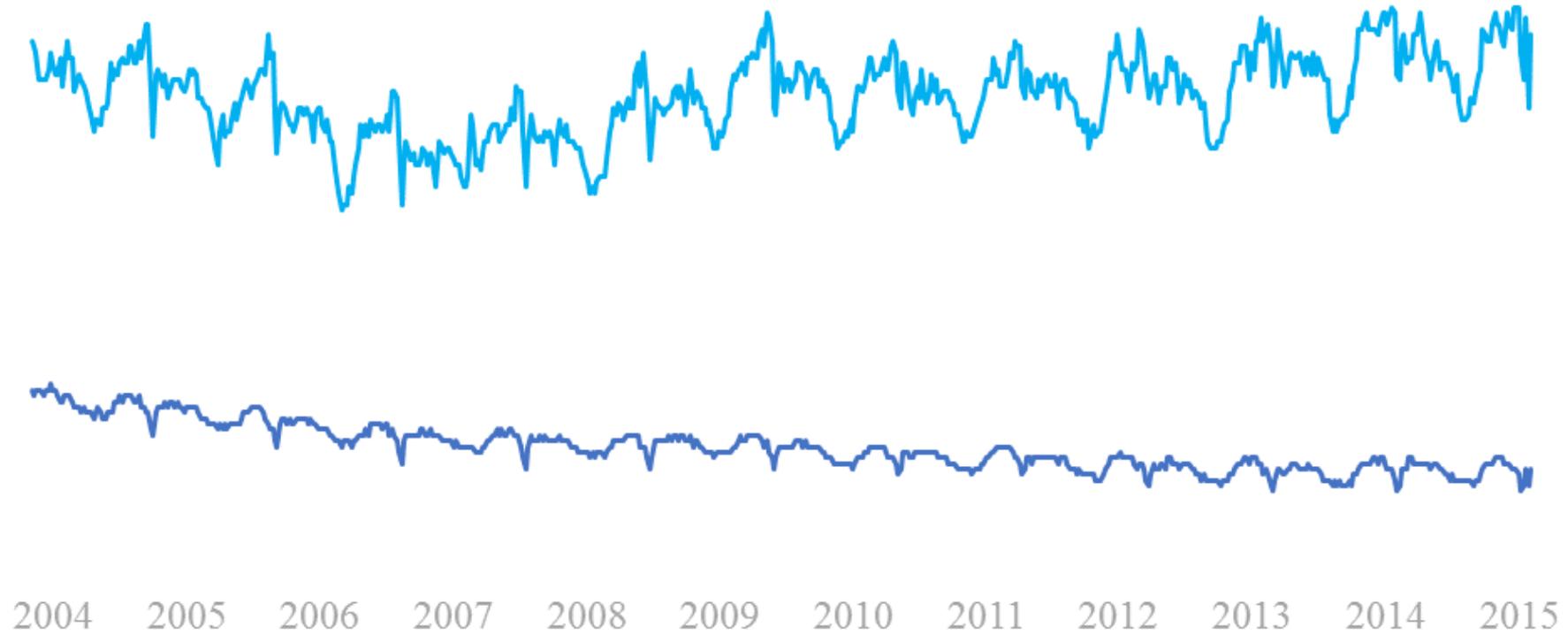
1. A site's pages are being crawled and effectively indexed by search engines
2. Pages are not only being indexed, but rising high enough to be visible in relevant SERP

My focus is on visibility for library resources, specifically the resources listed in Online Public Access Catalogues (OPACs)



Google trends data

Library Book





My research questions

1. What factors motivate libraries to address search engine visibility issues?
2. What barriers stand between Canadian public library information resources and search engine visibility?
3. What factors have the potential to enable libraries to achieve search engine visibility for their resources?



Methodology

- Semi-structured interviews with library staff members
- Site-specific searches of libraries' catalogue websites to see how many resources are currently being indexed by search engines
- Website traffic data collected by libraries about how users find their resources

How visible are Canadian public library resources right now?



 = Libraries that block crawlers from accessing the OPAC using robots.txt

Library	Number of Indexed OPAC Pages
1	481
2	214,000
3	269
4	29,100
5	51,700
6	7,370,000
7	221,000
8	207,000
9	213,000
10	1,410

[Web](#)[Images](#)[News](#)[Maps](#)[More ▾](#)[Search tools](#)

About 269 results (0.15 seconds)

Google promotion

✔ Try Google Search Console

www.google.com/webmasters/

Do you own [.net](#)? Get indexing and ranking data from Google.

? [.net/](#)

A description for this result is not available because of this site's robots.txt – learn more.

? [Explore the Catalogue](#)

[.net/client/en_US/default](#)

A description for this result is not available because of this site's robots.txt – learn more.

? [Catalogue](#)

[.net/client/en_US/default](#)

A description for this result is not available because of this site's robots.txt – learn more.

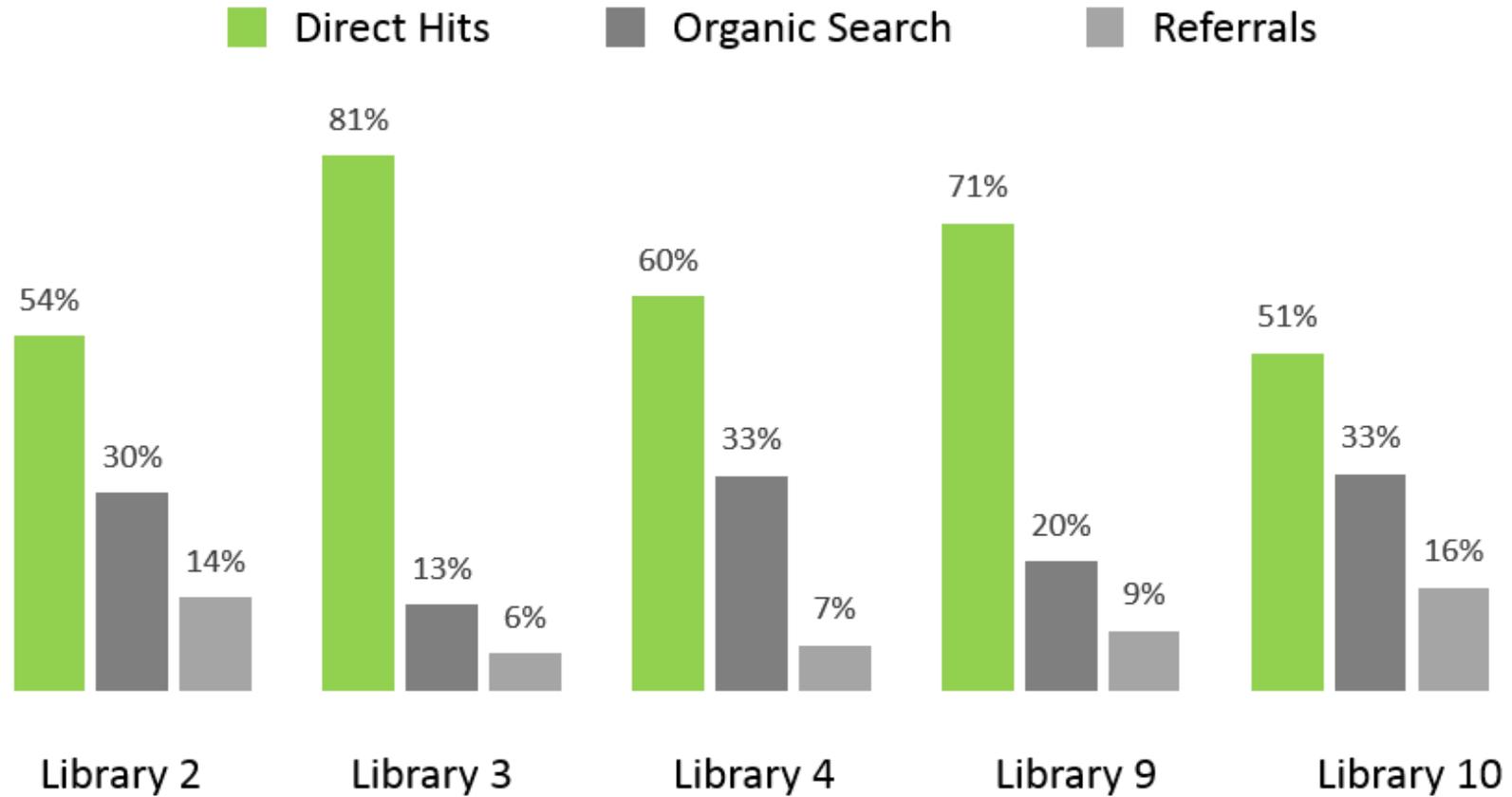
? [Explore the Catalogue](#)

[.net/client/en_US/default/](#)

A description for this result is not available because of this site's robots.txt – learn more.

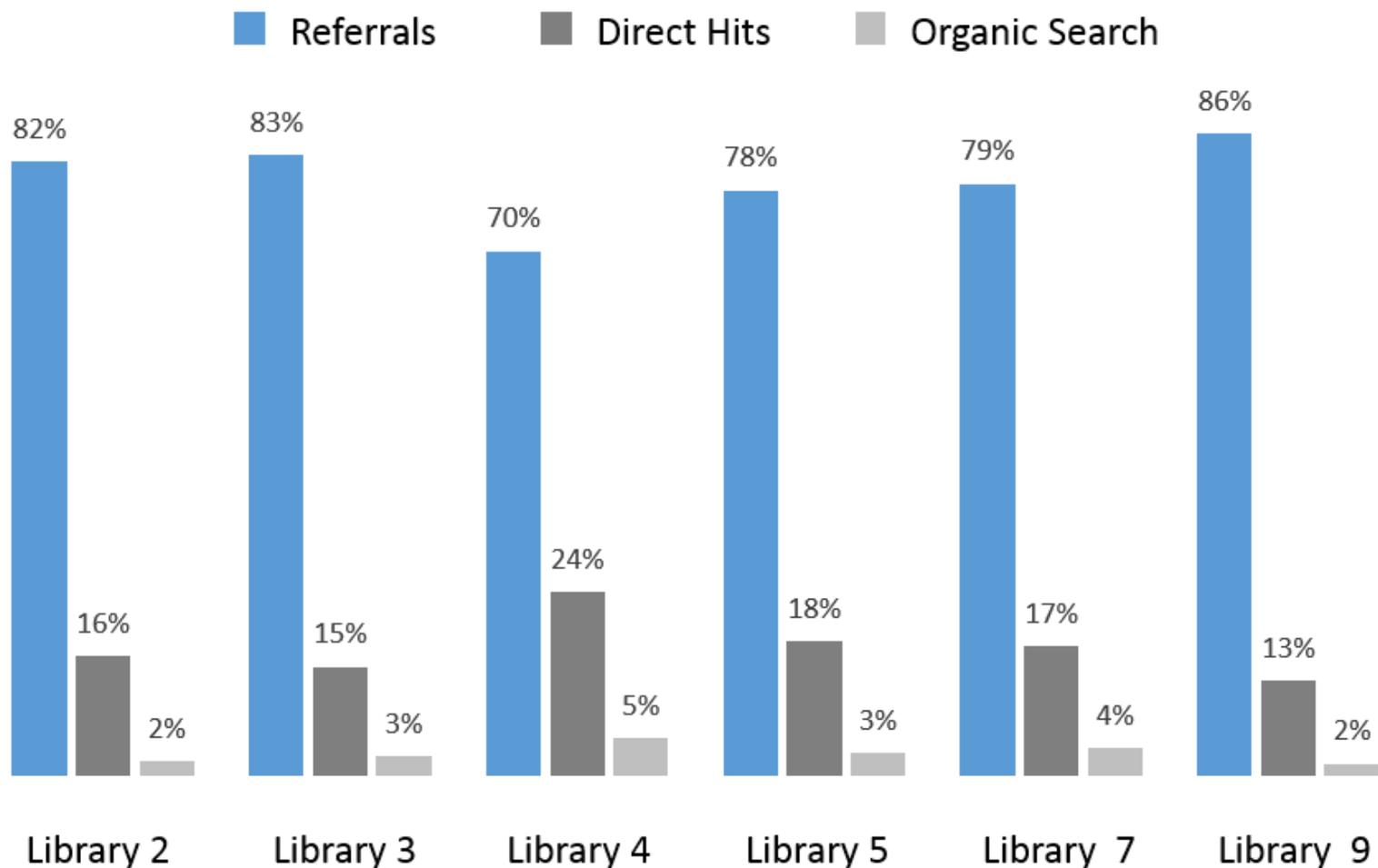


Website Traffic to Main Websites



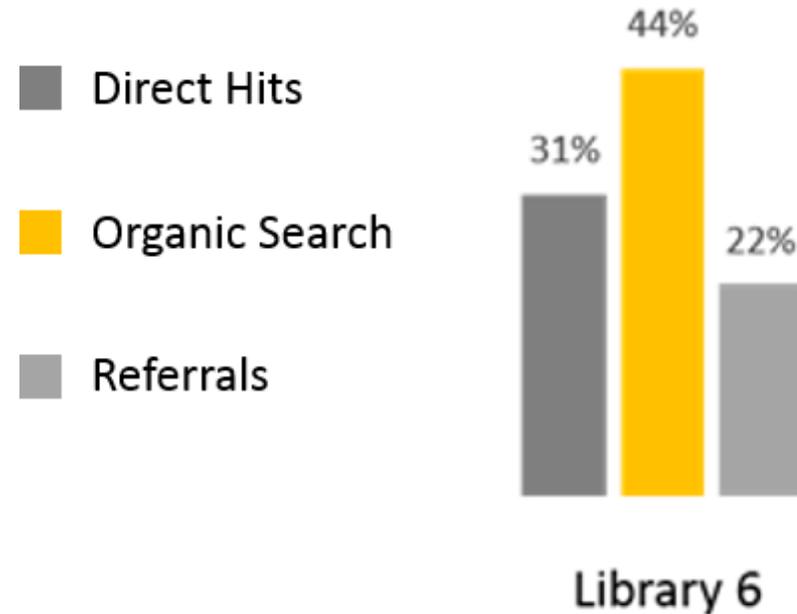


Website Traffic to OPAC Domains



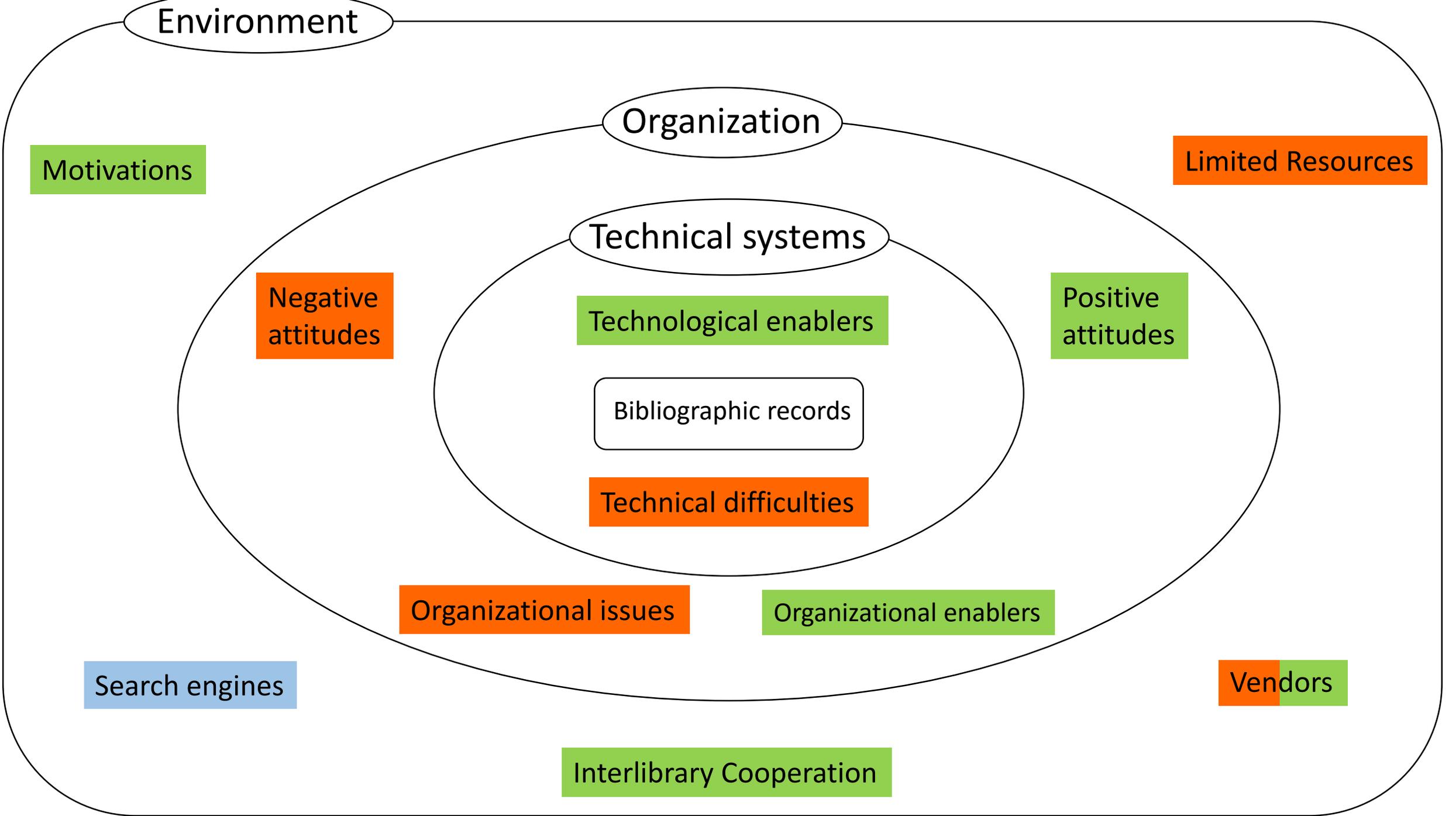


Library 6 - Unified OPAC and Website



L6: Top thousand search engine queries

Type of Query	Percent of the top thousand queries
Queries including L6's name	79.73%
Queries including general library keywords	15.71%
Queries including the name of a general resource (database, ebooks, audiobooks, etc)	3.28%
Queries including the name of a specific resource	0.76%
Other	0.51%



Environment

Organization

Technical systems

Motivations

Limited Resources

Negative attitudes

Technological enablers

Positive attitudes

Bibliographic records

Technical difficulties

Organizational issues

Organizational enablers

Search engines

Vendors

Interlibrary Cooperation



Vendors as enablers and barriers

- OPAC vendors often control crucial aspects of SEO for resources, and can be both enablers and barriers
- Some participants felt confident that over time, their vendors would provide improved SEO for library resources – others were skeptical
- Iain Lowe (Director of Product at Bibliocommons) has been experimenting with XML sitemaps and schema.org microdata
- He just did a webinar on the semantic web and SEO (<https://youtu.be/m9xITfgY77Y>)

Bibliocommons libraries before and after implementing XML sitemaps

Number of OPAC pages indexed in Google		
Library	First site-specific search (2015)	Second site-specific search (2016)
2	214,000	372,000
7	221,000	719,000
8	207,000	245,000
9	213,000	486,000



Technical Barriers: Difficulty being indexed

- Crawlers are blocked by robots.txt files
- Many databases still use query-based dynamic URLs
 - Search engines have evolved and can now interpret dynamic URLs, but not everyone is aware of this (Stiller & Szymanski, 2008)
 - Some databases use truly ephemeral URLs that are still a real barrier
- Desire to provide not simply stable URLs but *permanent* URLs before being indexed, in order to avoid 404 errors
- Lack of links to bibliographic records, due to structure of OPAC



Solutions

- Many of these problems can be solved by changing robots.txt files and creating XML sitemaps
- Robots.txt files tell crawlers what NOT to index
- XML sitemaps provide crawlers with a list of pages that SHOULD be indexed
- Dan Scott (2015) has a great step-by-step introduction to the use of robots.txt and XML sitemaps here:
<http://dx.doi.org/10.21083/partnership.v10i1.3328>



Technical Barriers: Rising to the top of SERP

- Even if all the resources in a catalogue are indexed by search engines, they may not rise high enough in search results to be seen
- For public libraries' location-specific resources, geolocation is important and doesn't seem to be currently applied to searches for individual book titles
- Many library catalogues do not conform to SEO best practices (e.g. having canonical URLs and using appropriate microdata), making them less attractive to search engines



Solutions

- Some problems can be solved by consulting the webmaster guidelines published by search engines
- For example, if a database uses faceted search, this can create multiple URLs for the same content. Search engines don't like this, but it's an easy fix:
 - Canonical URLs can be specified in a sitemap, or by adding a "rel=canonical" link element to duplicate webpages' HTML
(<https://support.google.com/webmasters/answer/139066>)
- For library content to rise high in SERP, search engines must be able to understand that content – this means content must be not simply machine readable, but machine **understandable**
- Basically, this adds up to linked data and the semantic web



Linked data: the Libhub Initiative

- Zepheira has recently emerged as a vendor of linked data solutions for library catalogues, with a strong emphasis on making library resources more visible online
- Their Libhub Initiative is essentially a test-case for LoC's BIBFRAME linked data model
- Right now, Libhub creates a separate linked-data version of the catalogue, which links back to each OPAC resource
- The idea is that by linking different libraries' bibliographic records, the libraries' combined online reputations will help each record rise higher in SERP



Linked Data: Schema.org

- The major search engines have developed a standardized structured data vocabulary: www.schema.org
- Learning and using this markup can solve many of the technical difficulties identified by participants
- Schema.org is not as granular as libraries would like, but the Schema Bib Extend group has succeeded in adding specialized vocabulary and other library-related modifications, making Schema.org markup more compatible with library needs (Scott, 2014)



Open questions

- How are search engines going to interpret libraries' SEO efforts?
- How can we make sure location-based information is being considered in searches for individual resources?
- If a library does everything right, will its resources actually reach the top of SERP?
- Is this even something that we should be attempting?



More info

- My thesis is available at <http://www.collectionscanada.gc.ca/obj/thesescanada/vol2/NSHD/T-C-NSHD-71421.pdf>
- You're welcome to get in touch with me at zoe.dickinson@dal.ca if you want to discuss any of this further

Suggestions for further reading

- Arlitsch, K. (2014a). Being irrelevant: How library data interchange standards have kept us off the internet. *Journal of Library Administration*, 54(7), 609-619. doi: 10.1080/01930826.2014.964031
- Arlitsch, K. (2014b). Semantic identity for library organizations. *Council on Library and Information Resources*. Retrieved from <http://connect.clir.org/blogs/kenning-arlitsch/2014/12/18/semantic-identity-for-library-organizations>
- Arlitsch, K., O'Brien, P., & Rossmann, B. (2013). Managing search engine optimization: An introduction for library administrators. *Journal of Library Administration*, 53(2-3), 177-188. doi: 10.1080/01930826.2013.853499
- Blandford, A. (2015). Google, public libraries, and the deep web. *Dalhousie Journal of Interdisciplinary Management*, 11. doi: 10.5931/djim.v11.1.5525
- Boston, T. (September, 2005). *Exposing the deep web to increase access to library collections*. Paper presented at The Eleventh Australasian World Wide Web Conference, Gold Coast, Australia. Retrieved from <http://ausweb.scu.edu.au/aw05/papers/refereed/boston/paper.html>
- Breeding, M. (2006). Systems librarian: How we funnelled searchers from Google to our collections by catering to web crawlers. *Computers In Libraries*, 26(4), 22-25.
- Breeding, M. (2014). The systems librarian: Enhancing discovery through better exposure. *Computers In Libraries*, 34(8), 24-26.
- Dickinson, Z., & Smit, M. (2015). Being where the people are: The challenges and benefits of search engine visibility for public libraries. *Library Hi Tech News*, 32(10), 11-15. doi: <http://dx.doi.org/10.1108/LHTN-08-2015-0055>
- Fichter, D., & Wisniewski, J. D. (2014). Being findable: Search engine optimization for library websites. *Online Searcher*, 38(5), 74-76.
- Onaifo, D., & Rasmussen, D. (2013). Increasing libraries' content findability on the web with search engine optimization. *Library Hi Tech*, 31(1), 87-108. doi: 10.1108/07378831311303958
- Scott, Dan. (2015). White hat search engine optimization (SEO): Structured web data for libraries. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 10(1). doi: <http://dx.doi.org/10.21083/partnership.v10i1.3328>
- Scott, Dan. (2014). Seeding structured data by default via open source library systems. *European Semantic Web Conference (ESWC) 2014*. Retrieved from <https://zone.biblio.laurentian.ca/dspace/handle/10219/2178>
- Thurrow, S. (2015). To optimize search, optimize the searcher. *Online Searcher*, 39(4), 44-48.
- Yang, L. (2016). Metadata effectiveness in internet discovery: An analysis of digital collection metadata Elements and Internet Search Engine Keywords. *College & Research Libraries*, 77(1), 7-19. doi:10.5860/crl.77.1.7

References

- DeRosa, C., Cantrell, J., Carlson, M., Gallagher, P., Hawk, J., & Sturtz, C. (2011). *Perceptions of libraries, 2010: Context and community*. OCLC, Inc. Retrieved November 23, 2014 from <http://oclc.org/reports/2010perceptions.en.html>
- Google. (2016). *Google Trends: Web search interest: Book, library: Worldwide, 2004-present* [dataset]. Retrieved from http://www.Google.com/trends/explore#q=%2Fm%2F0bt_c3%2C%20%2Fm%2F04h8h&cmpt=q&tz=Etc%2FGMT%2B4
- Lowe, Iain. (2016, June 17). *Improving visibility of libraries and library collections on the greater web: Bibliocommons update* [Video file]. Retrieved from <https://youtu.be/m9xITfgY77Y>
- Purcell, K., Brenner, J., & Rainie L. (2012, March 9). Search engine use 2012. Retrieved from <http://www.pewinternet.org/2012/03/09/main-findings-11/>
- Scheeren, W. O. (2012). *The hidden web: A sourcebook*. Santa Barbara, California: Libraries Unlimited.
- Scott, Dan. (2015). White hat search engine optimization (SEO): Structured web data for libraries. *Partnership: The Canadian Journal of Library and Information Practice and Research*, 10(1). doi: <http://dx.doi.org/10.21083/partnership.v10i1.3328>
- Scott, Dan. (2014). Seeding structured data by default via open source library systems. *European Semantic Web Conference (ESWC) 2014*. Retrieved from <https://zone.biblio.laurentian.ca/dspace/handle/10219/2178>
- Sherman, C., & Price, G. (2003). The invisible web: Uncovering sources search engines can't see. *Library Trends*, 52(2), 282-298. Retrieved from <http://hdl.handle.net/2142/8528>