Being where the people are: The challenges and benefits of search engine visibility for public libraries.

## Abstract

### Purpose:

The purpose of this article is to examine the challenges and benefits presented by search engine visibility for public libraries. The article outlines the preliminary results of a pilot study investigating search engine visibility in two Canadian public libraries, and discusses practical approaches to search engine visibility.

### Methodology:

The study consists of semi-structured interviews with librarians from two multi-branch Canadian public library systems, combined with quantitative data provided by each library, as well as data obtained through site-specific searches in Google and Bing. Possible barriers to visibility are identified through thematic analysis of the interviews. Practical approaches are identified by the author based on a literature review.

### Findings:

The initial findings of this pilot study identify a complex combination of barriers to visibility on search engines, in the form of attitudes, policies, organizational structures, and technological difficulties.

### Research limitations

This article describes a small, preliminary pilot study. More research is needed before any firm conclusions can be reached.

### Practical implications:

A review of the literature shows the increasing importance of search engine visibility for public libraries. This article outlines practical approaches which can be undertaken immediately by libraries, as well as delving into the underlying issues which may be affecting libraries' progress on the issue.

### Originality/Value:

There has been little original research investigating the reasons behind libraries' lack of visibility in search engine results pages. This article provides insight into a previously unexplored area by exploring public libraries' relationships with search engines.

## Libraries and Search Engines

As the online and offline realms become more and more interconnected, online information increasingly supports offline accessibility. This changes the way people expect to locate and access resources. Libraries remain important sources of information, but they are rarely the starting point for an information search. Instead, the vast majority of searches begin with online search engines.

According to a 2010 OCLC survey, 84% of online information seekers began their search using a search engine; 0% began with a library website (DeRosa et al., 2011). Furthermore, in a 2012 PEW Internet study, 91% of adult search engine users reported that they "always" or "almost always" found the information they were looking for using search engines; only 9% combined reported "only sometimes" or "hardly ever" (Purcell, Brenner & Rainie, 2012, p. 14). This shows not only widespread use, but widespread confidence in search engines as an information source.

## The Deep Web

Despite these positive public perceptions, search engines barely begin to encompass all the information available online. As of 2009, 95% of the web, representing over 220 billion pages, was not indexed by search engines (Scheeren, 2012). This is the deep web. The resources listed in libraries' Online Public Access Catalogues (OPACs) used to inhabit the deep web due to technical limitations, but as databases evolve and search engine crawlers improve, this hidden state is no longer a technological necessity.

Nevertheless, the contents of most Canadian public libraries' OPACs remain hidden in the deep web (Blandford, 2015). While most libraries have an online presence, few are visible online unless a user is specifically searching for a library. If a user simply searches for an information resource, the library remains invisible whether or not it offers the resource in question. By failing to integrate the library's information resources within search engines, they risk being invisible to the people who need their services.

There has been discussion of the relationship between libraries and search engines in the library and information science (LIS) literature, but less has been done to address the question of why most libraries' resources remain invisible to search engine users. If search engine visibility is technologically possible, why isn't it happening? This article will outline the preliminary results of a pilot study investigating this question, and suggest practical approaches to online visibility.

## LIS Perspectives

One possible explanation for libraries' continued invisibility can be found in a review of LIS literature. Many LIS scholars approach search engines as a threat. Some argue that the limitations of search engines make libraries more necessary than ever as champions of information literacy (e.g. Herring, 2005; Cahill & Chalut, 2009). Although these authors emphasize search engines' limited results, and discuss the deep web, there is no mention of the fact that this group of hidden resources often includes the library's own holdings. Others suggest that libraries should compete

by emulating the all-purpose usability of the Google search-box (e.g. Connaway & Randall, 2013). On the other hand, some say libraries should concede defeat and reposition themselves by offering different, more specialized services (e.g., Gorman, 2006). In all of these analyses, libraries are seen as reacting to the disruptive effects of search engines, rather than actively participating in this new information forum. This perspective may be dissuading library professionals from seeing search engine visibility as a possible or even desirable goal.

The last few years have seen the beginning of a movement to consider search engines as powerful tools, rather than threats. Proponents of this perspective advocate search engine optimization (SEO): the implementation of practices aimed at making resources friendly to search engine crawlers, in order to improve their visibility on search engine results pages (SERP).

Advocates for SEO argue that libraries must integrate their resources within the larger "information ecosystem" (Arlitsch, 2014, p. 610), and see search engines as a "funnel for channeling patrons back to the library" (Breeding, 2014, p. 26). From this perspective, becoming familiar with the latest SEO techniques for database content is essential (Breeding, 2014). In addition, change is necessary in both attitudes and departmental structures; rather than relegating SEO to the IT department, administrators must integrate SEO with their organization's overall mission in order to ensure staff at all levels are aware of its importance in reaching the community (Arlitsch, O'Brien, & Rossman, 2013).

The benefits of SEO for libraries may seem self-evident, but there is work to be done in proving that visibility in SERP would make a tangible difference for libraries. Measuring these possible effects is difficult because few libraries allow search engines to index their OPACs. However, a 2013 study of Ontario public libraries found "a strong statistically significant relationship between the numbers of library webpages indexed by Google and the number of users it receives" (Onaifo & Rasmussen, p. 102). This hints at the possibilities offered by SEO, but until more library content begins showing up in SERP, it is difficult to guess what impact that visibility might have. The literature to date shows that it is possible to make library resources compatible with search engines indexing and offers some arguments for doing so. However, few libraries have taken that step. What barriers stand between libraries and search engine visibility?

**Attitudes, Policies, and Structures as Barriers to Visibility**

The beginning of an answer to this question is provided by a pilot study conducted in 2015, consisting of semi-structured interviews with librarians in two major Canadian public library systems. Two types of quantitative data were also considered: first, the number of OPAC pages currently indexed by search engines was determined using site-specific searches in Google and Bing (see Table 1); second, the main sources of traffic to each OPAC were determined using website usage statistics provided by each library system (see Tables 2 and 3).

Being where the people are: The challenges and benefits of search engine visibility for public libraries.

**Table 1**. Total Indexed Pages

### Number of OPAC Pages Indexed

| Search Engine | Library 1 | Library 2 |
|---|---|---|
| Google | 1140 | 238,000 |
| Bing | 4870 | 37,000 |

**Table 2**. Library 1's most popular landing pages

### Most Popular Landing Pages

| URL | Percentage of Sessions |
|---|---|
| Main Site | 62.32% |
| OPAC | 7.78% |
| Main Site (Mobile Version) | 4.91% |
| Main Site Index | 2.03% |

**Table 3**. Source of Visitors for Library 2's OPAC and Main Site

| Source | OPAC | Main Site |
|---|---|---|
| Referral | 72% | 14% |
| Direct | 26% | 54% |
| Organic Search | 2% | 30% |
| Social | 0% | 1% |
| Email | 0% | 1% |
| (Other) | 0% | 0% |

There remains much to be done in investigating this issue. However, even this small pilot study provides insight into a previously unexplored area. Through thematic analysis, a complex combination of barriers was identified as standing between library resources and search engine visibility, including: attitudes, policies, organizational structures, and technical difficulties.

*Attitudes*

Participant 1 appears to be representative of the negative feelings about search engines outlined above. This librarian describes search engines as untrustworthy companies that "make millions of dollars off of essentially the free information of the world", and is wary of allowing them access to library data. While recognizing the importance of SEO in publicizing the library's main site, he would prefer to trust the data contained in the library's OPAC to more library-centric services such as WorldCat. This individual also describes a general lack of awareness of search engines in their

organization as a whole. These negative attitudes make surmounting any other barriers unlikely, because doing so is not seen as a desirable goal. Participant 2 has a positive attitude towards search engines, and reported predominantly positive attitudes within their organization.

*Policies*

Both participants' library systems do not have any official policies regarding SEO or interaction with search engines. In both cases, the participants themselves are in the process of building such a policy. Lack of a policy may be a barrier to search engine visibility; without consistent policies, any SEO initiatives which do arise may lack clarity, standards, and metrics for success.

*Structures*

Search engine visibility for library resources sometimes falls between the cracks because it is an interdepartmental issue. For example, in Participant 1's library, web development and SEO happen in a different department (and different library branch) from the development of the OPAC. Neither department has direct responsibility for the online visibility of the OPAC's resources, which makes it difficult to address the issue, or even become aware that such an issue exists. Participant 2's library has mitigated this by creating an interdepartmental taskforce with members from three different departments. However, Participant 2 still encounters structural difficulties because the library's OPAC is provided by an outside vendor, and that vendor retains control over the OPAC's interaction with the rest of the web. SEO initiatives cannot simply be implemented by the library's taskforce; instead, a dialogue must be opened with the vendor, significantly slowing the process.

**Technical Difficulties**

It is now technologically possible to open up library OPAC contents to search engines. However, doing so is not necessarily easy. In the past, difficulty of indexing OPACs arose largely because of dynamic URLs: the content of the database was displayed in pages created in response to each specific query. This could cause crawlers to get lost in mazes of almost-identical temporary pages, or be unable to advance beyond the initial query page. Some OPACs still have dynamic URLs, but many newer products offer static URLs: that is, each resource has a URL which persists whether or not it has been queried. While static URLs remain preferable, crawler technology has improved to the point of being able to interpret and index dynamic URLs (Stiller & Szymanski, 2008). Technically, search engines now have the capacity to crawl any OPAC. However, in many cases, including Library 1, crawlers are blocked from entering the OPAC. Even though some URLs from that domain are indexed by search engines (see Table 1), they are essentially blank because the site's Robots.txt file does not allow crawlers access to any information (see Figure 1).

Being where the people are: The challenges and benefits of search engine visibility for public libraries.

**Figure 1**. Library 1's Search Results
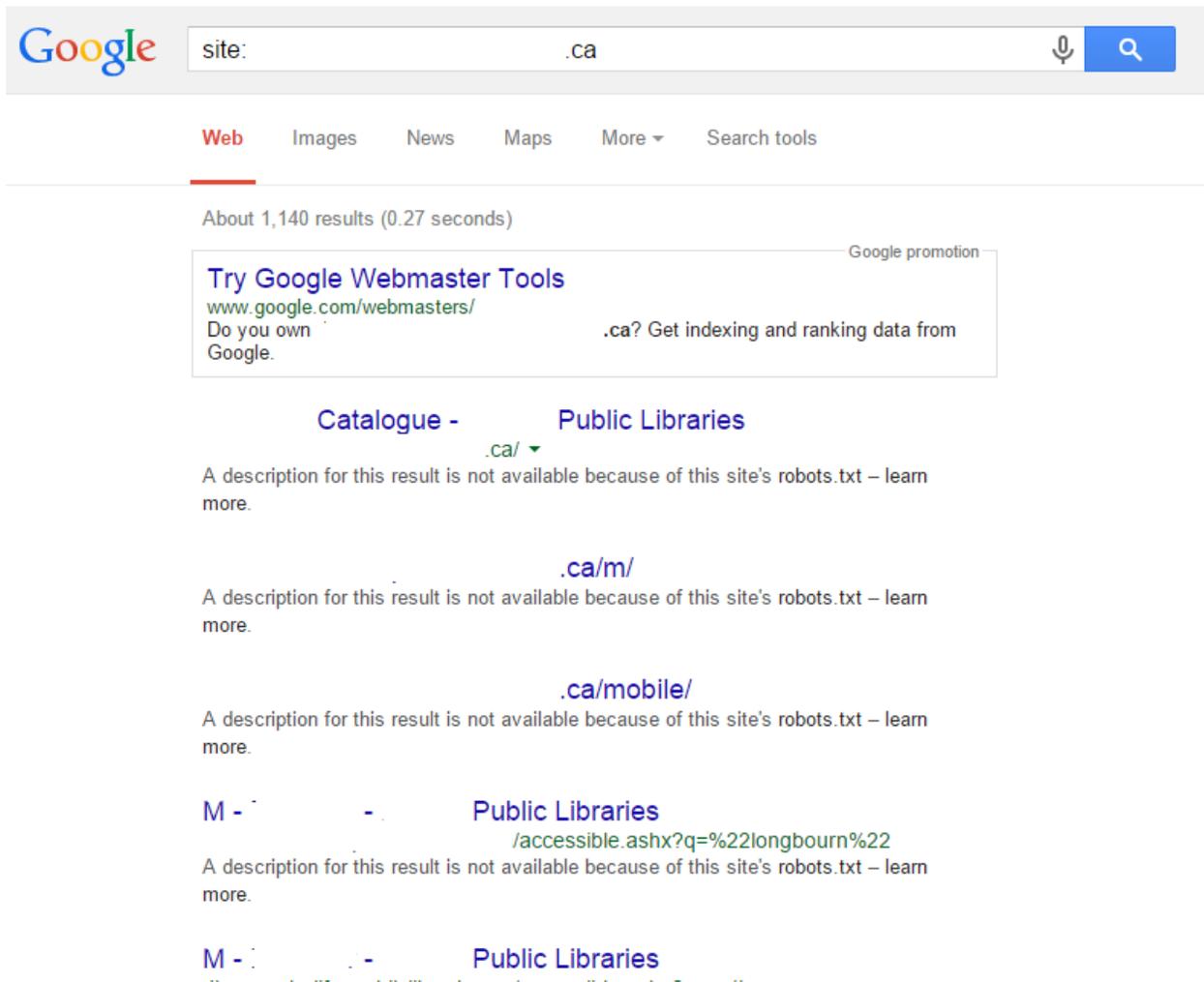


Library 2, on the other hand, has static URLS and over 200,000 OPAC pages indexed (see Table 1). Its Robots.txt allows crawlers, and the results are rich in metadata (see Figure 2).

**Figure 2**. Library 2's Search Results
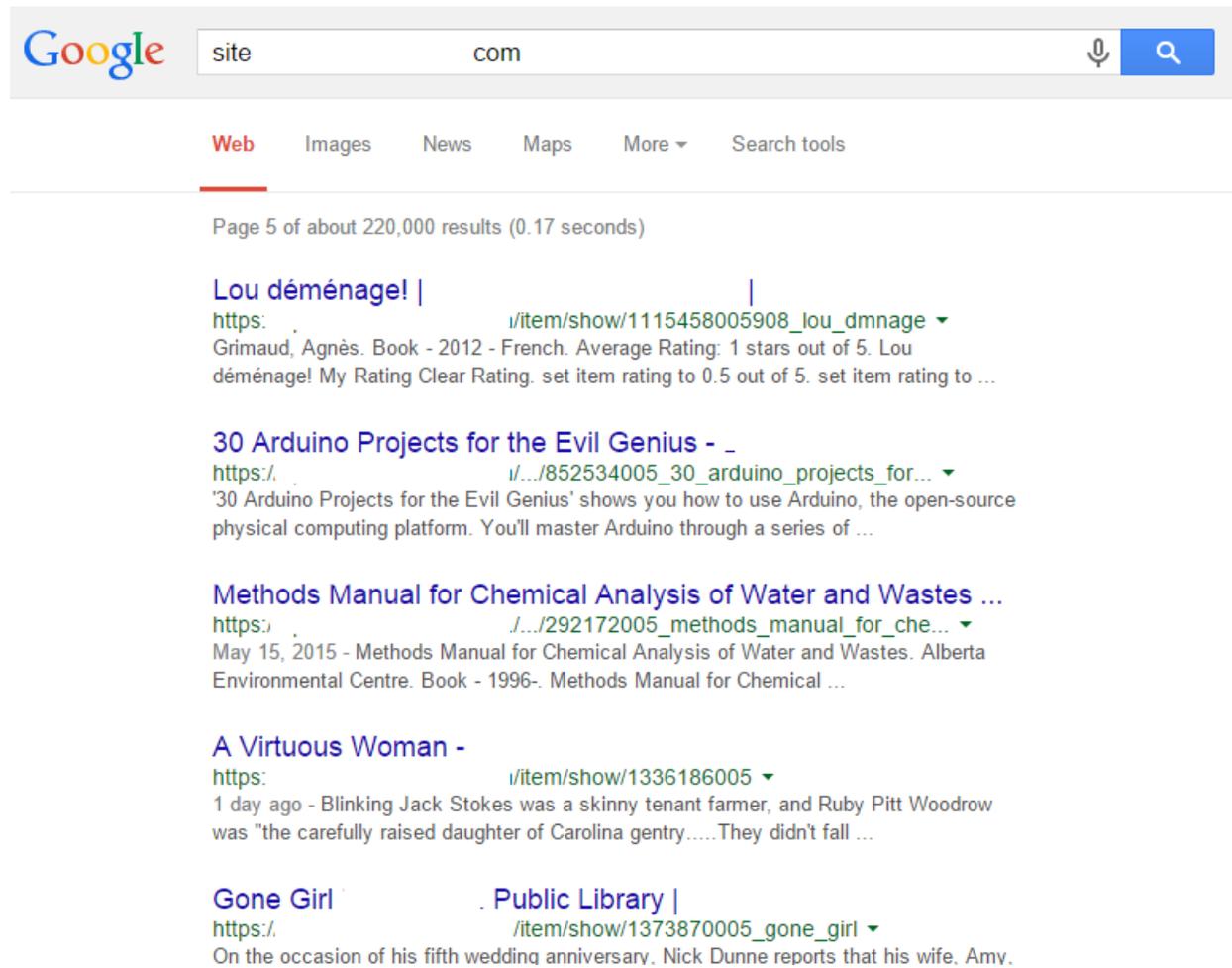


However, being indexed does not immediately equate to visibility. Resources must show up high enough in SERP to be truly visible to searchers. This is where Participant 2's team has run into difficulties. In preliminary investigations, it was found that: if you just had geolocation turned on and you did a basic search like, "Angie Abdou Bone Cage", we're not showing up in the top five SERP. However, if you type in the word "X Town," which I thought was bizarre because I explicitly turned on my geolocation which is X Town, then we do show up in the top, quite often first SERP.

While this level of visibility is an improvement over none at all, it must be balanced against the unlikelihood of searchers including their location when searching for a book. This may account for the fact that while Participant 2's library system has almost half of its OPAC content indexed by Google, only about 2% of the web traffic experienced by its OPAC is coming from search engines.

Another technical barrier reported by Participant 2 was difficulty in maintaining not only static, but *permanent* URLs for resources. This means making sure that each item's URL persists not only from one search to the next, but from one year to the next, no matter how many times items

come and go from the catalog. According to Participant 2, libraries are working to achieve this first, because "prioritizing site indexing and pushing our way up the SERPs before providing permanent URLs will point users in the wrong direction and potentially result in frustration or mistrust."

It is interesting to note, however, that while maintaining permanent URLs may be difficult, it is not generally seen as a necessary prerequisite for being crawled by search engines. Content providers simply expect search engines to crawl their pages frequently in order to correct any URL changes which have taken place. The changeability of the Internet is widely acknowledged, and an occasional 404-error is par for the course. Participant 2's desire to conquer this problem before prioritizing site indexing lends credence to Arlitsch's (2014) suggestion that librarians' conscientious attachment to high data standards may be hindering their ability to mesh with the rest of the internet.

## Practical Approaches

The findings of this pilot study suggest that the first step in improving the visibility of a library's resources in SERP is to make interoperability with search engines a goal. This may mean addressing negative attitudes about search engines, and shifting expectations about what is required in order to open the library's resources to indexing. This may also mean reshaping organizational structures, building interdepartmental taskforces, writing policies, and generally speaking integrating the goal of search engine visibility into the institution's overall mission. Until these shifts in attitude and structure take place, it is difficult for an organization to approach the issue. Given the control many OPAC vendors maintain over catalogues' online interactions, opening a dialogue with these organizations may also be an important step.

It may take time to make the organizational changes necessary to truly integrate search engine visibility with a library's strategic goals. However, there are practical steps that libraries can take at any time to begin opening up their OPACs so search engine indexing.

### Robots Exclusion Protocol (http://www.robotstxt.org/)

The Robots Exclusion Protocol is a method for controlling how much of a domain is indexed by search engine crawlers, and how frequently. Making sure an OPAC's Robots.txt file correctly directs crawlers is the first step in making library resources visible. Sometimes the OPAC vendor, rather than the library, controls the OPAC's robots.txt file. However, anyone can examine a website's robots.txt file, by typing the domain name followed by /robots.txt (e.g., http://domain-name.com/robots.txt). Whether or not a library has control over the file, it can be examined by library staff and its content can be discussed with the OPAC vendor.

### Sitemaps (http://www.sitemaps.org/)

The Sitemaps protocol offers search engine crawlers a comprehensive list of the URLs which the library wishes to have indexed. Creating a sitemap is a relatively straightforward way to make sure the appropriate pages are being crawled.

## Linked Data

### *Schema.org*

Schema.org is a collaborative initiative to build structured data standards online. Using Schema.org's standardized microdata, webmasters can make it easier for machines such as crawlers to understand the content and meaning of web pages (Breeding, 2014). This makes it easier for search engines to index sites, and connect searchers with relevant materials. Schema.org is sponsored by all the major search engines, making it a good bet for communicating with crawlers. Compared to libraries' meticulous metadata, Schema.org's all-purpose markup does lack granularity, overlooking concepts that are important to libraries (Enis, 2015). However, a group of librarians and stakeholders, known as Schema Bib Extend (see https://www.w3.org/community/schemabibex/), has banded together to help extend Schema.org's ability to represent bibliographic information (Enis, 2015). In the meantime, Schema.org microdata can be added to OPAC pages, alongside more traditional, library-centred metadata.

### *The LibHub Initiative* (http://www.libhub.org/)

MARC cataloging may still be the current reality for libraries, but the LibHub initiative seeks to step out of that paradigm and actively explore the possibilities of linked data for library visibility. LibHub participants are essentially early adopters of the Library of Congress' BIBFRAME linked data standards, test-driving linked data protocols with their catalogues (Zepheira, 2014). Joining this initiative is a step libraries can take, not only in becoming more visible online, but also in integrating that goal with the rest of the library's functions.

Schema Bib Extend and BIBFRAME may seem to be working at cross-purposes, in that they are independently developing two different sets of linked data standards for libraries. However, a recent joint paper co-published by the Library of Congress and the OCLC suggests that eventually, the two initiatives will complement each other (Godby & Denenberg, 2015). For an overview of current linked data projects in libraries, see the OCLC's recent survey of linked data implementers (Smith-Yoshimura, 2014).

## Conclusion

As Participant 2 stated, "As a public librarian […] you want to be where the people are". This means having a presence in the most popular information sources online: search engines. They may not be perfect, but they are an essential tool in remaining accessible to patrons. This pilot study has merely scratched the surface of this complex and pressing issue. More research is necessary into the challenges faced by libraries in becoming more visible, and the possible impacts of such visibility. Further research on this topic has the potential to help libraries compare notes, learn from each-others' challenges and triumphs, and better understand the barriers and opportunities inherent in online visibility initiatives.

## References

Arlitsch, K. (2014), "Being irrelevant: How library data interchange standards have kept us off the internet", *Journal of Library Administration,* Vol. 54, No. 7, pp. 609-619, doi: 10.1080/01930826.2014.964031

Arlitsch, K., O'Brien, P., & Rossmann, B. (2013), "Managing search engine optimization: An introduction for library administrators." *Journal of Library Administration*, Vol. 53, No. 2-3, pp. 177-188, doi: 10.1080/01930826.2013.853499

Blandford, A. (2015), "Google, public libraries, and the deep web", *Dalhousie Journal of Interdisciplinary Management,* Vol. 11, doi: 10.5931/djim.v11.1.5525

Breeding, M. (2014), "The systems librarian: Enhancing discovery through better exposure", *Computers In Libraries,* Vol. 34, No. 8, pp. 24-26.

Cahill, K. & Chalut, R. (2009), "Optimal results: What libraries need to know about Google and search engine optimization", *The Reference Librarian*, Vol. 50, No. 3, pp. 234-247, doi: 10.1080/02763870902961969.

Connaway, L. S. & Randall, K. M. (2013), "Why the internet is more attractive than the library", *The Serials Librarian,* Vol. 64, pp. 41–56, doi: 10.1080/0361526X.2013.761053.

DeRosa, C., Cantrell, J., Carlson, M., Gallagher, P., Hawk, J., & Sturtz, C, (2011), "Perceptions of libraries, 2010: Context and community", OCLC, Inc, available at: oclc.org/reports/2010perceptions.en.html (accessed 23 November, 2014).

Enis, M. (2015), "OCLC works toward linked data environment - ALA midwinter 2015", *Library Journal.* Available at: lj.libraryjournal.com/2015/02/technology/oclc-works-toward-linked-data-environment-ala-midwinter-2015 (accessed 3 July, 2015).

Godby, C. J., & Denenberg, R. (2015), *Common ground: Exploring compatibilities between the linked data models of the Library of Congress and OCLC.* Dublin, Ohio: Library of Congress and OCLC Research. Available at:

www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf (accessed 3 July, 2015).

Gorman, G.E. (2006), "Giving way to Google", *Online Information Review,* Vol. 30, No. 2, pp. 97-99, doi: 10.1108/14684520610659148.

Onaifo, D., & Rasmussen, D. (2013), "Increasing libraries' content findability on the web with search engine optimization", *Library Hi Tech,* Vol. 31, No. 1, pp. 87-108, doi: 10.1108/07378831311303958.

Purcell, K., Brenner, J., & Rainie L. (2012), "Search engine use 2012", available at: http://www.pewinternet.org/2012/03/09/main-findings-11/ (accessed 3 February, 2015).

Scheeren, W. O. (2012), *The Hidden Web: A Sourcebook*, Libraries Unlimited, Santa Barbara, California.

Being where the people are: The challenges and benefits of search engine visibility for public libraries.

Smith-Yoshimura, K (2014), "Linked data survey results 1 – Who's doing it", available at http://hangingtogether.org/?p=4137 (accessed 23 July, 2015).

Stiller, J. & Szymanski, K. (2008), "Dynamic URLS vs static URLs", available at http://googlewebmastercentral.blogspot.ca/2008/09/dynamic-urls-vs-static-urls.html (accessed 31 July, 2015).

Zepheira, LLC. (2014), "Join the movement – take the Libhub Initiative pledge", available at: http://www.libhub.org (accessed March 15, 2015).