PUBLIC LIBRARIES AND SEARCH ENGINES

by

Zoe Dickinson

Submitted in partial fulfilment of the requirements
for the degree of Master of Library and Information Studies

at

Dalhousie University
Halifax, Nova Scotia
April 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Although public libraries are dedicated to connecting people to information, the information resources offered by many Canadian libraries are currently inaccessible to search engine indexing robots, and thus effectively invisible to online searchers. This mixed-method study investigates factors that may be impacting the visibility of Canadian public library resources on search engines through semi-structured interviews with library staff in ten libraries across Canada. Quantitative data about each library's website is used to provide a more complete picture of the current visibility of participating libraries. The study identifies attitudes, organizational factors, and technological factors that may be motivating, inhibiting, and enabling search engine visibility, and provides recommendations on how to address these factors. By delving into the underlying factors which may be affecting libraries' progress on the issue, this study aims to help inform libraries' decision-making processes and practices, as well as clarifying avenues for further research.

LIST OF ABBREVIATIONS USED

HTML	Hypertext Markup Language, a language for creating web pages.
HTTP	Hypertext Transfer Protocol, a protocol for transmitting data online
ILS	Integrated Library System
IT	Information Technology
LIS	Library and Information Science
MARC	Machine Readable Cataloguing
OCLC	Online Computer Library Centre
OPAC	Online Public Access Catalogue
SEO	Search Engine Optimization
SERP	Search Engine Results Page
URL	Uniform Resource Locator

GLOSSARY

BIBFRAME: Bibliographic Framework, a data model for bibliographic description. BIBFRAME was designed to replace MARC standards, and to use linked data principles to make bibliographic data more useful both within and outside the library community.

Bing: A search engine promoted by Microsoft Inc.

Canonical URLs: A way to inform search engines of the preferred URL in a set of URLs with duplicate content.

Crawler: A program that visits Websites and reads their pages and other information in order to create entries for a search engine index.

Discovery Layer: A search interface used by libraries to allow patrons to explore library collections, including third-party content such as vendor-provided e-books, in a user-friendly way.

Domain name: The component of a URL that identifies a specific website (e.g., www.**domain.com**/subpage).

Faceted Search: A search interface that allows users to explore a collection of information by applying multiple filters.

Google: A search engine widely used on the World Wide Web.

Google Analytics: A service offered by Google that tracks website traffic.

Hyperlink: An electronic link that, when clicked, directs the user to another webpage, or another part of the same webpage.

Knowledge Graph: Images and information provided in SERP by Google in an attempt to answer searchers' questions directly, using Semantic Web technology.

Libhub: The Libhub Initiative aims to raise the visibility of libraries on the web by actively exploring the promise of BIBFRAME and Linked Data.

Linked Data: The building blocks of the Semantic Web; a method of connecting data on the Web from different sources.

MARC: A data model developed in the 1960's to represent bibliographic data in a machine-readable format.

Metadata: Data that provides information about other data.

Microdata: A set of tags used in HTML to help search engines and other applications better understand the content of webpages.

Organic search: In website traffic data recorded about the sources of visitors to a website, organic search is one possible source of hits, referring to visitors that reached the website in question by clicking on a link in a search engine results page. This does not include hits originating from paid advertisements on search engines.

Rich Snippet: Descriptive data displayed in SERP by Google from appropriately structured web pages.

Robots Exclusion Protocol: A standard used by websites to communicate with web crawlers and other web robots, involving the use of a /robots.txt file to specify acceptable crawling behaviour for a specific domain.

Schema.org: Provides a collection of shared vocabularies webmasters can use to mark up their pages in ways that can be understood by the major search engines.

Semantic Web: An extension of the existing World Wide Web. It provides a standardized way of expressing the relationships between web pages and semantic concepts, to allow machines to understand the meaning of hyperlinked information.

Sitemap: A hierarchical list of webpages used to guide crawlers or users through a website.

Surfacing Content: Allowing search engines to index online content in order to make it accessible to searchers.

ACKNOWLEDGEMENTS

Many thanks to the library staff members across Canada who took the time to speak with me during the course of this study. I sincerely enjoyed speaking with each of you.

A huge thank you to my thesis advisory committee: my supervisor, Dr. Mike Smit, who is a mentor in the truest sense of the word, and fosters curiosity and rigorous inquiry in all of his students; Dr. Louise Spiteri, whose depth of expertise and attention to detail have vastly improved this document; and Dr. Sandra Toze, whose insightful and constructive feedback helped me tell the story of my research with clarity and conviction. Thank you all for taking so much time to guide me through this adventure. Many thanks also to my external reader, Dr. Ali Shiri, for his kind and thorough assessment of my work.

Thank you to Kim and Patricia Dickinson, my tireless non-specialist readers, for their unconditional support every step of the way. Thank you to Chris Brown, for his encouragement, support, and faith in my abilities. Thanks to Chantel Ridsdale for helping me sit down and write that dreaded first page, and to Alieda Blandford for the long-ago kitchen-table conversation that first sparked this line of inquiry.

CHAPTER 1 – INTRODUCTION

The rise of the internet to its current ubiquity has changed the way people think about their information needs. Not only are there more potential sources for information, but searchers use different techniques to identify those sources. Search engines have become the first stop in the hunt for information. Whether or not search engines are the best possible source for any given piece of information, they are almost always the starting place for identifying the best sources. According to a 2010 Online Computer Library Center (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.

Despite these positive public perceptions, search engines barely begin to encompass all the information available online. As of 2009, an estimated 95% of the web, representing over 220 billion pages, was not indexed by search engines (Scheeren, 2012). This hidden content is known as the deep web. Since then, the web has grown so large that such estimations are no longer possible (BrightPlanet, 2014). Despite improvement in search engines' ability to index database content, vast quantities of information, including the high quality content contained in the databases of memory institutions, such as museums, archives, and libraries, remain concealed in the deep web (Sherman & Price, 2003; Scheeren, 2012; Cahill & Chalut, 2009). This situation impedes the average person's access to information; not only is the quality of retrieved information limited (even though the quantity may be large), users may not realize the search results are incomplete, and thus have no incentive to dig deeper.

While most libraries do have an online presence, few are visible in search engine results pages (SERP) unless a user is specifically searching for a library. If a user simply

searches for an information resource, the library remains invisible even if it offers the resource in question. Libraries may still be the best possible source for certain services and information, but the resources offered by many Canadian libraries are not visible in SERP (Blandford, 2015). If libraries do not integrate their information resources with search engines, they risk being invisible to the people who need their services. This endangers not only the perceived relevance of libraries, but also the average searcher's ability to access quality information. Public access to information is essential to the well-being of both individuals and society (Bishop, 2012). If public libraries, as champions for public information access, allow themselves to drop out of sight in today's most popular information forum, Canadians may not be aware of the essential services they offer.

Library resources have traditionally been excluded from search engine indexes due to the technological limitations of their Online Public Access Catalogues (OPACs). OPACs provide online listings of physical books contained in libraries' collections, digital resources such as e-books and audiobooks, and sometimes even listings of library programming. In their earlier forms, OPACs could only be searched through portals provided on library websites. Over the last ten years, OPAC technology has evolved, and the indexing ability of search engines has improved. This hidden state is no longer a technological necessity. Nevertheless, many Canadian public library catalogues remain hidden in the deep web (Blandford, 2015).

1.1 Purpose of the Research

As the online and offline realms become more and more interconnected, online information increasingly supports offline accessibility. A recent report from the Council of Canadian Academies (2015) acknowledges the importance of online visibility for memory institutions such as libraries:

> In all facets of our lives, we expect citizen-centric services to seamlessly interact with how we use and access digital material and information every day. If documentary heritage is to be used in the shaping of Canada's culture, it must be digitally discoverable and accessible. (p. xv)

Similarly, the New Media Consortium's *NMC Horizon Report: 2015 Library Edition* identifies "competition from alternate avenues of discovery" as a "difficult challenge" for libraries (Johnson, Adams Becker, Estrada & Freeman, 2015, p.26), and warns that "emerging Internet technologies are fostering changes in patron behavior, challenging libraries to either adapt to the new expectations defined by current discovery practices or risk becoming obsolete over time" (p. 26). Keeping pace with the public's changing information seeking habits can be an ongoing challenge for memory institutions. Canadian public libraries offer essential resources and services to their communities, but have fallen behind in making those resources truly accessible and discoverable online.

The purpose of this research is to investigate factors that impact Canadian public libraries' ability to become more visible on search engines, in order to answer the following three research questions:

RQ1: What factors motivate libraries to address search engine visibility issues?
RQ2: What barriers stand between Canadian public library information resources and search engine visibility?
RQ3: What factors have the potential to enable libraries to achieve search engine visibility for their resources?

This study is the first to investigate Canadian public libraries' current relationships with search engines. There has been discussion of the relationship between libraries and search engines in library and information science (LIS) literature (e.g., Blandford, 2015; Arlitsch, 2014a; Breeding, 2014), but less has been done to address the question of why so many libraries' resources remain invisible to search engine users. If search engine visibility is technologically possible, what factors stand between Canadian public library information resources and search engine visibility? What factors motivate, inhibit, and enable librarians to improve their libraries' visibility? Because these research questions center around libraries' action or inaction when it comes to search engine visibility, the first step in providing answers is to discuss the issue of search engine visibility with library staff members. Through semi-structured interviews with library staff members

across Canada, this study identifies possible motivators, barriers, and enablers to making library resources visible on search engines. This study also assesses the current visibility of each library's information resources by examining quantitative data in the form of site-specific searches and website usage statistics. This quantitative data provides an important point of comparison to aid in the analysis of qualitative data. The data gathered here provides a starting point for understanding and improving Canadian public libraries' relationship with search engines.

1.2 An Overview of Relevant Technology

1.2.1 The Mechanics of Library Catalogue Technology

The software used to display library resources online can impact the visibility of library resources in SERP. Libraries use a complex array of technology to manage and display library resources. In most cases, this technology is provided by a third party, or vendor. Multiple software and hardware components operate in varying configurations, depending on the needs of the library in question, in order to manage and display bibliographic records, programming information, and digital resources. Vendors of library technology are perceived as having an important role in either enabling or discouraging visibility initiatives (see Section 4.3.5 and Section 4.4.5). This means that understanding the technological components libraries often purchase from vendors is important. This brief overview of the technology involved in managing library resources will provide a foundation for further discussion of the interactions between libraries, vendors, and search engines.

Figure 1 (below) shows a visual representation of the different layers of technology that may be used by libraries to manage bibliographic records and their appearance online, placed within the larger context of the Internet as a whole. Every library does not use every component shown in Figure 1, and libraries can use these components in different ways depending on their needs. Search engine crawlers are shown unconnected to any other elements. Depending on the permissions contained in each element's /robots.txt

file, search engine crawlers may interact with the information contained in various different components shown here.

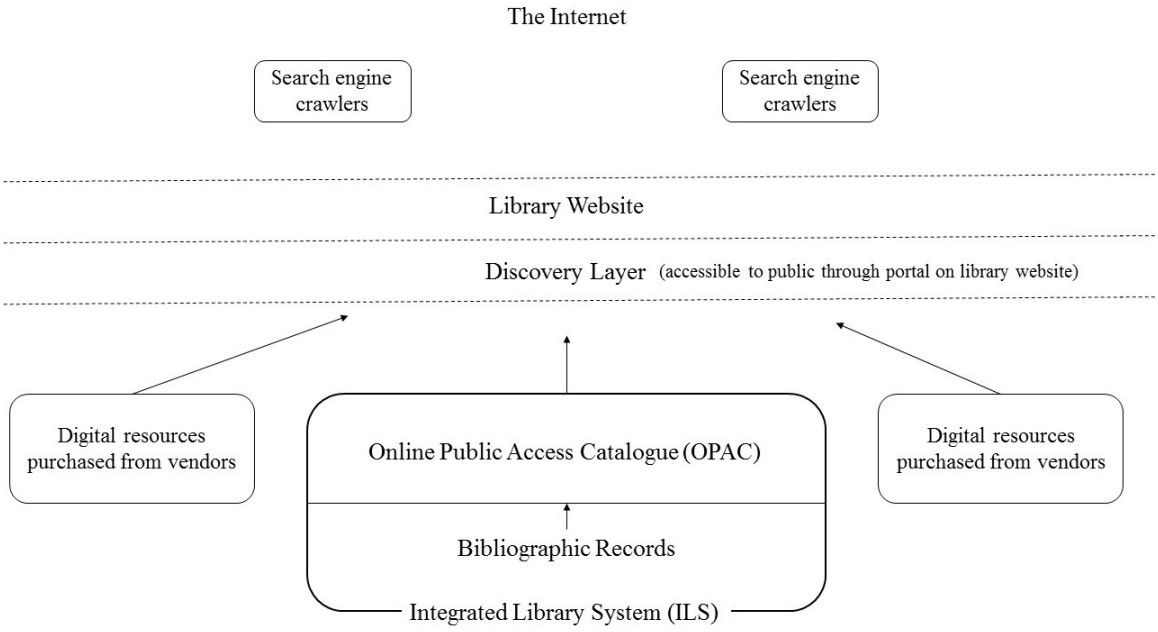


Figure 1: An overview of technology used by public libraries to manage and display bibliographic records

The core technology used to manage library resources is the integrated library system, or ILS. An ILS is "an integrated set of applications designed to perform the business and technical functions of a library, including acquisitions, cataloging, circulation, and the provision of public access" (Reitz, 2014a). For the purposes of this study, the most relevant component of the ILS is the online public access catalogue, or OPAC. An OPAC is "a database composed of bibliographic records describing the books and other materials owned by a library or library system" (Reitz, 2014b). This catalogue, and the records it contains, must be crawled and interpreted by search engines in order to make library resources visible in SERP.

Most libraries also offer their patrons digital resources such as e-books, audiobooks, digital magazines, and digital music. These resources are obtained from vendors such as Overdrive (www.overdrive.com), Zinio (www.zinio.com), and Naxos Music Library (www.naxosmusiclibrary.com), which often come with their own interfaces, databases, and search functions. As library resources become increasingly diverse, traditional

OPACs have struggled to provide a user-friendly search experience that encompasses the whole library collection, including digital resources from other vendors, rather than being limited to catalogue records (Christison, 2013). Due to this gap, an added piece of technology has come into common usage (Christison, 2013). The discovery layer or discovery service is a search interface "built around unified indexes of licensed scholarly publications combined with metadata for the local catalogue and other local content, with a Google-like simple search box, and a consolidated, single result list" (Christison, 2013, p. 2). In recent years, some OPAC vendors have begun to provide next-generation catalogues that attempt to provide a similar search experience to discovery layers without need of an added layer of technology (Christison, 2013).

For the purposes of this study, the distinction between an OPAC and a discovery layer is only peripherally relevant. It is important to note this distinction, and understand the different layers of technology that allow libraries to interact with users online. However, the focus of this research is on the online visibility of library content in general, rather than the mechanics of managing library content. In this study, the term OPAC is generally used to refer to libraries' online catalogues, with the understanding that in some cases there is also a discovery layer in place, which may also impact the way library resources are interpreted by search engines.

1.2.2 The Mechanics of Search Engines

Understanding the current visibility of library resources in SERP also requires a solid understanding of how search engines work. Search engines are an integral part of daily life for many people, but the mechanics of their operation may not always be well understood. Throughout this document, Google is often referred to as an example, both by myself and by participants, in discussing search engines. As the most commonly used search engine by far, it is not surprising that Google has become synonymous with search (Purcell, Brenner & Rainie, 2012). However, most of the information presented here applies to other search engines as well.

The web is incredibly vast, and constantly growing; at last count, Google's index of webpages used "well over 100,000,000 gigabytes" of storage (Google, n.d.a, para. 1), and, as noted in Section 1.1, there remain vast amounts of information not included in that index. The web is too big, too chaotic, and too dynamic to be organized by human experts. Google and other search engines have taken on the task of indexing this vast and disorganized collection of pages, using robots called crawlers to gather information from each page.

Web crawlers are automated software programs that travel the web, adding webpages to search engines' indexes. They begin crawling with a list of webpages based on previous crawls, as well as "sitemaps provided by website owners" (Google, n.d.a, para. 3). In order to discover new content, crawlers follow hyperlinks from known webpages, moving from site to site until they run out of links to follow. The information gathered by crawlers about each domain they visit is then added to the search engine's index, which is constantly being updated. The frequency of web crawler visits and the number of pages visited for each website is determined by computer programs, and can also be controlled by webmasters using the Robots Exclusion Protocol (Google, n.d.a).

Once pages are added to search engines' indexes, they can be retrieved in response to a searcher's query. Search results are provided in search engine results pages (SERP), which list links to relevant pages in order of relevance. A decision about which pages to provide in answer to a specific query is made using a complex set of computer algorithms in approximately one-eighth of a second (Google, n.d.b). Algorithms are used to analyze each query and determine its intent, as well as to determine which pages in the search engine's index will be most relevant to the searcher (Google, n.d.b). Google's most famous algorithm is the PageRank algorithm, which judges the quality of a website based on how many other sites link to it, and on the reputation of those linking sites. This algorithm allows Google to make sense of the web by tapping into millions of website creators' decisions about other websites' quality; if many high-quality pages link to a particular webpage, it too is likely to be high-quality (Levy, 2011). However, PageRank is only one of many factors. Over the years, the process has become extremely complex,

and increasingly personalized. Google considers over 200 different factors in deciding which pages to retrieve in answer to a given query (Google, n.d.b). As search becomes more user-centred, many of those factors have become personalized, based on the user's past search behaviour, location, language, and what device is being used to make the search (SEOBook, 2016).

There are many ways in which website owners can attempt to control the way search engines index and retrieve their content. The following sections briefly outline some of these techniques.

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

With the Robots Exclusion Protocol, website owners can use a /robots.txt file to request that crawlers stop indexing a site altogether, disallow crawlers from specific pages or sections of a site, exclude specific robots from indexing a site, and control how frequently crawlers should access a site ("About /robots.txt", 2007). There are some robots crawling the web that do not respect these requests; however, crawlers used by the major search engines, such as Google and Bing, will abide by the Robots Exclusion Protocol ("About /robots.txt", 2007). Making sure an OPAC's /robots.txt file correctly directs crawlers is the first step in making library resources visible. Often 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 this file, it can be examined by library staff and its content can be discussed with the OPAC vendor.

1.2.2.2 The Sitemaps Protocol (http://www.sitemaps.org/)

The Sitemaps Protocol offers search engine crawlers a comprehensive, hierarchical list of the URLs that a website owner wishes to have indexed. Creating a sitemap is a relatively straightforward way to make sure the appropriate pages are being crawled.

1.2.2.3 Canonical URLs

In some cases, particularly when information resources are managed by complex databases, the same content can be accessed through many different URLs. Many duplicate URLs can be problematic for search engines, as this forces them to index and manage multiple versions of the same content. This can negatively impact a page's rank in search results. Google's recommended solution for this problem is to specify a canonical URL for content that is available through multiple URLs. This tells crawlers which version of the URL to add to the search engine's index. Canonical URLs can be specified in a sitemap, or by adding a "rel=canonical" link element to duplicate webpages' HTML specifying which version is the canonical version. Google offers guidelines for defining canonical URLs in the Search Console Help Center (Google Inc, 2016c).

1.2.2.4 Schema.org Markup (www.schema.org)

Schema.org is a collaborative initiative to build structured data standards online. Using schema.org's standardized microdata, website owners can make it easier for search engine crawlers to understand the content and meaning of web pages (Breeding, 2014). Microdata is a set of tags used in HTML to help search engines and other applications better understand the content of webpages ("Why use microdata", n.d.). This makes it easier for search engines to index sites, and connect searchers with relevant materials. Using appropriate schema.org markup helps search engines identify appropriate content for rich snippets in search results, among other benefits. Schema.org is sponsored by all the major search engines, making it a good bet for communicating with crawlers. Schema.org microdata can be added to OPAC pages, alongside more traditional, library-centred metadata. For more information about schema.org microdata as it relates to the Semantic Web and linked data, see Section 2.2.4.

1.3 Organization of the Thesis

This thesis is organized into six chapters. Following this introductory chapter, Chapter 2 provides a review of approaches to search engines found in LIS literature, as well as an

overview of the latest developments in search engine technology and the semantic web. Chapter 3 outlines the methodology used throughout the research project. Chapter 4 presents the results of the interviews, as well as reporting on the quantitative data gathered for each library. Chapter 5 provides an analysis of the motivators, barriers, and enablers identified in Chapter 4, and discusses relationships between themes. Chapter 6 sums up this study's original contribution to the LIS field, offers recommendations to libraries based on an analysis of the data, and suggests avenues for further research.

CHAPTER 2 – LITERATURE REVIEW

An extensive review of literature was performed, including searches in library science databases such as Library, Information Science & Technology Abstract (LISTA), Library Literature & Information Science full text (LLIS), computer science databases such as ACM Digital Library, and multidisciplinary databases such as Google Scholar, ProQuest, and EBSCOhost. This literature review provides an overview of existing literature on the topic of search engine visibility in the context of library science. Two contrasting approaches to search engines in the LIS field are described in Section 2.1, ranging from negative perceptions of search engines as threats to libraries and library values, to positive perceptions of search engines as potential outreach tools. In Section 2.2, the Semantic Web and linked data are presented as an alternative approach to online information seeking, and their potential to address search engine visibility for libraries is discussed. Gaps in existing literature are discussed in Section 2.3, and this study is proposed as a first step in addressing those gaps.

2.1 Approaches to Search Engines in the LIS Field

2.1.1 Search Engines as the Competition

In her 2015 article, Blandford (2015) suggests one possible explanation for libraries' continued invisibility on search engines: "librarians tend to see Google as a rival and a threat" (p. 13). A review of LIS literature seems to support this suggestion. Many LIS scholars approach search engines as a threat to libraries' relevance. Responses to search engines as a perceived threat have been varied within the library community. Some have dismissed the idea that search engines could be anything more than an additional tool in the librarian's expert research arsenal (Norris, 2006; Herring, 2005). Others suggest that there is no contest; Google has already won, and libraries should reposition themselves by offering different, more specialized services (e.g., Gorman, 2006).

Due to the high quantity and uncertain quality of the sources provided by search engines, librarians involved in reference service tend to see search engines as an information literacy challenge, and urge their colleagues to embrace the role of information guide rather than information provider (e.g., Cahill & Chalut, 2009; Egger-Sider & Devine, 2005). Many of these analyses include discussions of the deep web, and emphasize the fact that search engines provide limited results. Cahill and Chalut (2009), for instance, note that "there is a wealth of information on the deep web locked behind barriers, such as registration and authentication, that users who never look beyond Google are missing" (245). There is a keen awareness that patrons are missing important information by relying solely on search engines. Librarians are presented as offering valuable supplementary resources and information literacy education; however, there is no mention of the fact that this group of hidden resources often includes the library catalogue itself. How will patrons ever connect with this valuable service, if the library which offers it has only limited presence on the very search engines in question?

At the same time as some librarians have been urging patrons and colleagues to be critical of search engines' simplicity (Cahill & Chalut, 2009; Egger-Sider & Devine, 2005), there has been a movement to mimic that simplicity in the OPAC itself. There is increasing demand for library systems that can respond to patrons' evolving usability standards. Numerous studies have shown that users' experiences with search engines influence their expectations when using other information retrieval tools such as OPACs (e.g., Kumar, 2011; Griffiths & Brophy, 2005). Connaway and Randall examine twelve different user behaviour studies and conclude that libraries "need to make our systems look more like search engines" (2013, p. 52). Similarly, Lim's (2009) survey of both traditional and next generation OPACs shows newer products seek to mirror Google's one-stop-shop mentality by widening the scope of search within OPACs and even integrating broader resources such as Google Book Search and WorldCat.org into their systems. For further discussion of traditional and next generation OPACs, please see Section 1.2.1.

A recent Wall Street Journal op-ed piece, and the debate it sparked, confirms that the narrative of Google as a threat to libraries is still alive and well. Barker (2016), a librarian himself, argues that thanks to the internet, and Google specifically, professional librarians are "a dying breed". The article received 55 comments in about two weeks, with strong opinions on both sides of the debate. Many librarians, including the president of the American Library Association, wrote to defend their relevance, citing libraries' importance in information literacy education and bridging the digital divide, as well as arguing that the LIS profession is successfully adapting to changing technology ("Librarians' role changes as information does", 2016).

Both information literacy initiatives and attempts to match search engines' intuitive usability are valuable responses to the issues raised by search engines. However, in treating search engines as competition, both approaches miss an important point. In all of these analyses, libraries are seen as reacting to the disruptive effects of search engines, rather than actively participating in this information forum. Based on the popularity of search engines among information seekers (see, for example, Purcell, Brenner & Rainie, 2012), it is apparent that search engines are not competition for libraries; they are the information ecosystem within which their patrons live. In order to be useful to the public they have served for so long, libraries must also live in that ecosystem. This means seeing search engines not as a threat, but as a critical outreach tool for communicating with both local and worldwide patrons.

2.1.2 Search Engines as Outreach Tools

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 in SERP. Advocates for SEO argue that increased visibility in SERP is an essential step in reaching out to library patrons. Blandford (2015) goes so far as to suggest that failure to do so will "contribute

[…] to libraries' perceived irrelevance in the digital age" (p. 2). From this perspective, becoming familiar with the latest SEO techniques for database content is essential.

Kenning Arlitsch, along with various colleagues, argues for increased institutional awareness of SEO within libraries. Arlitsch spent years working on digital library repositories, only to find that they were hardly being used; in the course of investigating that issue, he became an enthusiastic proponent of SEO (Arlitsch, 2014a). According to Arlitsch, libraries' focus on complex and rich metadata standards that do not translate well into the wider online world has not only been labour intensive and costly, but has actively inhibited interoperability between libraries and the rest of the web (2014). He argues that to remain relevant, libraries must make their content visible in the larger "information ecosystem", rather than forcing customers to move to different, library-specific systems (Arlitsch, 2014a, p. 610). Approaching this problem as an administrator, his advice is to stop relegating SEO to IT departments; rather, administrators must integrate SEO with an organization's overall mission and make sure staff at all levels are aware of its importance in reaching the community (Arlitsch, O'Brien & Rossmann, 2013).

Thurow (2015) also emphasizes SEO's importance as an outreach strategy, arguing that "the beneficiary and target of SEO techniques are not only search engines. The ultimate target and beneficiary are searchers" (p. 44). She identifies negative perceptions of search engine optimization, where SEO is seen as exploitative or dishonest, as a possible factor dissuading librarians from investing time and energy into SEO, and advocates increased education and training on the topic in order to counteract these negative stereotypes. Thurow (2015) argues that becoming adept in SEO techniques is a necessary part of librarians' mandate as information professionals: "SEO responsibilities should fall in the hands of people who have optimized documents for findability for many, many centuries: librarians" (p. 46). According to Thurow (2015), optimizing library content for search engine findability is not only important in helping libraries promote their content, but also in ensuring adequate access to information in their communities: "I urge search

professionals, librarians, and information scientists to consider where their skills and expertise can contribute to a more educated online population" (p. 48).

Marshall Breeding, a regular columnist for *Computers in Libraries* magazine, also argues that libraries must make their resources "findable on the open web and in the broader universe of information" (2014, p. 24). In contrast to Arlitsch's and Thurow's more administrative approach, Breeding attacks the problem from a technical perspective. He provides detailed advice on both basic and semantic web SEO techniques, including example coding and directions to important online resources such as sitemaps.org. He also directs readers to schema.org, home of the standardized, structured-data markup system preferred by major search engines. While emphasizing the continued importance of the OPAC as a specialized discovery tool, Breeding advocates for SEO techniques as a "funnel for channeling patrons back to the library" (26). Similarly, Fichter and Wisniewski (2014) offer practical techniques for navigating the technology involved in SEO, and outline SEO best practices from a technical point of view. They remind librarians that "Since for many users, the search engines are the front door to all that your site has to offer, investing in SEO pays great dividends" (p. 76). However, their advice centres on the library's main website, rather than library resources.

From a more technical perspective, Yang's 2016 study seeks to determine the true effectiveness of digital library metadata in enhancing discoverability of collections. Yang's analysis offers a quantitative assessment of the effectiveness of librarians' current metadata practices in facilitating retrieval through search engines. Yang (2016) found that the Dublin Core metadata fields dc.title, dc.description, and dc.subject "played the most significant roles in facilitating discovery of digital items by Internet search engines" (p. 17). This research is significant in offering research-based recommendations about where librarians should focus their efforts in trying to make library metadata more compatible with search engines. Yang's study was limited to one university's digital repository, and does not begin to address questions about metadata in a public library setting. However, the methodology used offers interesting possibilities for future research in that area. In this case, as with Arlitsch and Breeding, the conversation is specifically about digital

resources; however, much of this discussion would also be applicable to the physical collections represented in OPACs.

While many commercial business owners would consider the benefits of SEO obvious, this is not the case in the library community. As Fichter and Wisniewski (2014) note, "Being findable in search results is so important, many companies invest heavily in SEO efforts. Unlike companies, libraries usually lack the resources to hire SEO experts or to dedicate a team or person to SEO" (p. 74). If libraries are to invest their limited resources in SEO, it is essential to have clear evidence of SEO's importance in helping patrons access library resources. There has been some research which supports the utility of SEO for libraries. A 2013 study of Ontario public libraries found "a strong statistically significant relationship between the number of library webpages indexed by Google and the number of users [a website] receives" (Onaifo & Rasmussen, p. 102). Tony Boston of the National Library of Australia presents a more in-depth case study documenting the results of opening up a collection of digital images to search engines (2005). In 2002, the Pictures Catalogue was opened to search engine robots; by 2004, there was a 370% increase in the use of its collection (Boston, 2005). Marshal Breeding has also published a record of his own experiences using these techniques, detailing his successful use of sitemaps to make a digital repository indexable by search engines (2006). After allowing time for content to filter through Google's system, Breeding recorded not only an increase in traffic, but an increase in resource use: "from August 2005 to January 2006, roughly 30 percent of our videotape requests can be directly attributed to this path [from search engines to the newly indexed content]" (Breeding, 2006, p. 25).

These studies hint at the possibilities offered by SEO, but a true assessment of the impact of search engine visibility is hindered by the widespread lack of that visibility throughout the library community. Until more library content begins showing up in SERPs, it is difficult to predict the impact of this visibility.

2.2 The Semantic Web and Linked Data

2.2.1 What is the Semantic Web?

The Semantic Web is a developing concept that has the potential to revolutionize the way information is retrieved online. If this potential is realized, the Semantic Web could have huge implications for the visibility of library resources. As more and more library data is published in accordance with Semantic Web principles, the hope is that library data will become interoperable with the rest of the web, and as a result will be more visible and more effectively accessible to the average searcher.

The concept of the Semantic Web was introduced in 2001 by Tim Berners-Lee, one of the original architects of the web as we know it (Calaresu & Shiri, 2015). Berners-Lee envisioned the Semantic Web as the next step in the web's evolution, which would "bring structure to the meaningful content of Web pages," making it possible for machines to interpret the meaning of and understand the relationships between online information (Berners-Lee, Hendler, & Lassila, 2001, p. 36). The web consists of three main components: a vast store of information contained in linked documents; a vast community of human users, who both contribute to and consume that information; and the software applications that allow users to interact with online content (Calaresu & Shiri, 2015). However, the human users are the only component of that system who can truly make sense of online information.

Software applications such as search engines can be used to navigate the web and organize information, but those applications are "limited in their capacity to manipulate and use data embedded within the documents" (Calaresu & Shiri, 2015, p. 94). They cannot understand the semantic meaning of a piece of information, or the relationships between different pieces of information. For example, a human user reading a description of *Great Expectations* would identify Charles Dickens as the creator of the novel, and understand that Charles Dickens is an author who wrote many novels. A search engine robot would crawl a description of *Great Expectations* and retrieve a set of keywords, but gain no underlying understanding of the relationship between author and novel. As the web continues to grow (as of 2014, Google had indexed approximately 30 trillion pages

(Statistic Brain Research Institute, 2015)), this inability of machines to interpret data meaningfully becomes a serious challenge. As Calaresu and Shiri (2015) explain, the problem is that, "although the volume of information encoded into language and sent across the Web can only realistically be ordered by machines, human language itself is often far too complex and nuanced for any of our current machines to meaningfully interpret" (p.84). The more machines can understand not only the syntax but also the meaning, or semantics, behind content, the more efficient and precise automated functions such as information retrieval become, and the more possible it is to navigate the vast seas of information on the internet.

2.2.2 How Does Linked Data Fit In?

The terms "Semantic Web" and "linked data" are often used interchangeably, but they are not synonymous. As with the Semantic Web, definitions of linked data can vary, but one prevailing view states that linked data is the building block of the Semantic Web; "the Semantic Web is the whole, while Linked Data is the parts" (Heath, n.d.). The Semantic Web proposes to make information not simply machine-readable, but truly machine-understandable; linked data is a method for carrying out that intention. A discussion of the technical aspects of linked data is beyond the scope of this thesis. Suffice it to say that linked data uses Resource Description Framework (a data model used to describe information objects and their relationships to each other), Uniform Resource Identifiers (permanent and unique identifiers attached to information objects), and the Hypertext Transfer Protocol or HTTP (the standard information-exchange protocol used online) in order to express and transmit the semantic meaning of information objects in a way that computers can interpret (Heath, n.d.).

2.2.3 Semantic Search

The concept of the Semantic Web is still in its formative stages, with various different models being developed, and few practical applications in existence (Calaresu & Shiri, 2015). However, in recent years search engines Google and Bing have implemented Semantic Web technology in order to provide smarter, more intuitive search results (Uyar

& Aliyu, 2015). Google's Knowledge Graph and Bing's Satori both provide images and information in SERP in an attempt to answer searchers' questions directly, in addition to presenting them with a list of search results identifying relevant websites. Using semantic analysis to inform search represents a fundamental shift in search technology; rather than simply matching query text against indexed content and providing a list of possible matches, Knowledge Graph and Satori attempt to understand the meaning behind users' questions, and provide exact and appropriate answers (Uyar & Aliyu, 2015).

Since 2012, an increasing number of search results have used semantic search technology to include supplementary information on topics inferred from the keywords used. For example, in the search for "go set a watchman" below (Figure 2), Google displays the usual list of links on the left, based on the keywords used and their frequency and proximity to each other in the documents retrieved, as well as other factors built into Google's algorithms. On the right there is a knowledge panel: a square of images and text which responds to the query in an entirely different way. In order to display this panel, Google's Knowledge Graph system had to recognize that the keywords "go set a watchman", when taken together, correspond with a popular novel written by Harper Lee; instead of simply pulling documents from the web that match the keywords, Google understood the underlying meaning of the query, and provided relevant information, including some suggestions for further reading.

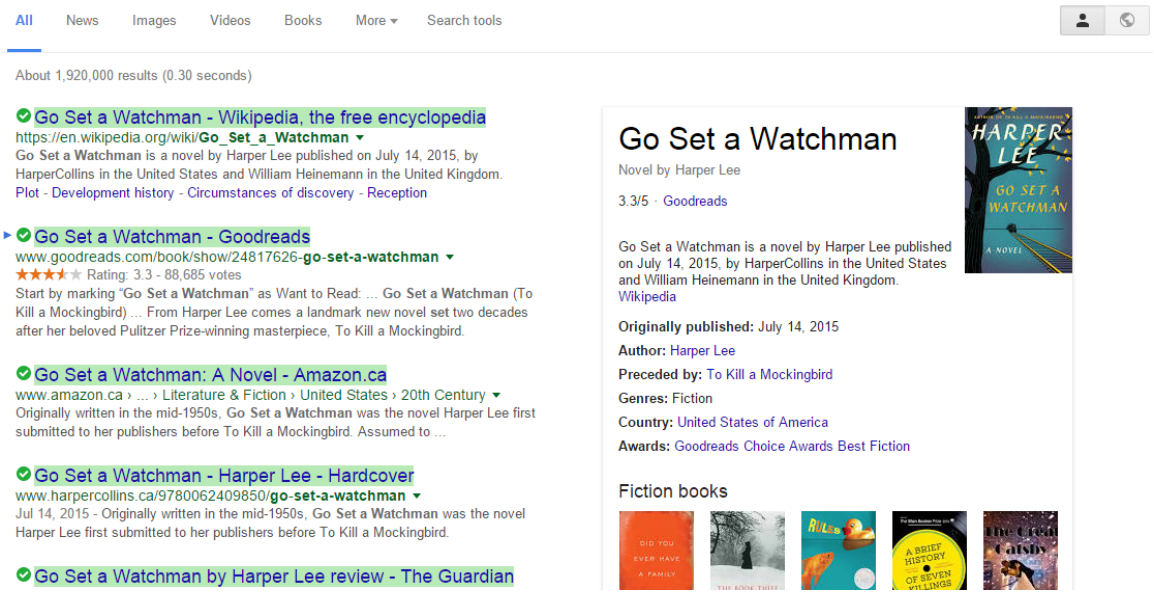


Figure 2: Google search results page showing Google Knowledge Graph information

2.2.4 Linked Data in Libraries

Moving beyond the more basic tenets of SEO (such as sitemaps), library linked data initiatives seek to build a new, internet-friendly foundation for library records. Most libraries use Machine Readable Cataloguing (MARC) format to contain data describing each resource in their catalogues. This format, which has been in use since the 1960s, is not easily parsed by search engine crawlers, and does not integrate well with the rest of the web (Enis, 2015a). Since 2011, the Library of Congress has been developing a linked data standard to fill this need: BIBFRAME (Enis, 2015a). In June 2014, the Libhub Initiative was launched as a test case for BIBFRAME (Zepheira LLC, 2014a). Working under the overarching pledge "I believe everyone benefits from the visibility of libraries and their content on the Web", the initiative is gathering early adopters to test the BIBFRAME linked data structure in working library environments (Zepheira LLC, 2014b). Of the ten libraries included in this study, two are Libhub early adopters.

While the Library of Congress has been working on BIBFRAME as an alternative to both MARC and schema.org's less detailed data structure, Schema Bib Extend has been working instead to expand and improve schema.org for library use. This group, which includes the Online Computer Library Centre (OCLC), argues that while schema.org's markup may not yet be granular enough to properly address bibliographic data, as the official chosen markup language of search engines such as Google and Bing, it should be considered the starting point for library visibility on the open web (Enis, 2015b).

These two initiatives may seem incompatible, but both parties hope that they will eventually complement each other. A joint paper co-published by the OCLC and the Library of Congress explores the differences in the two linked data structures, as well as ways in which the two projects are aligned (Godby & Denenberg, 2015). The projects do overlap, but schema.org remains "necessarily broad but shallow because library resources must compete with creative works offered by many other communities in the information

landscape", while BIBFRAME's coverage is "deep because it contains the vocabulary required of the next-generation standard for describing library collections" (Godby & Denenberg, 2015, p. 4). The hope of both initiatives is that they will eventually complement each other, with BIBFRAME serving as a new data model for libraries, and schema.org serving as a more generally-usable "data aggregator to import and export BIBFRAME data" (Godby & Denenberg, 2015, p. 8).

Another, more immediate step libraries can take to get involved in linked data is to participate in already-existing Semantic Web implementations. Arlitsch (2014b) observes that "library organizations are poorly represented in Semantic Web applications such as Knowledge Cards," and urges librarians to remedy that fact (para. 1). As discussed in Section 2.2.3, Google's Knowledge Graph, which is expressed in SERP through various manifestations including Knowledge Cards, uses Semantic Web technology to provide searchers with added information alongside search results (for an example of a Knowledge Card, please see Figure 2 on page 19). Librarians may be unsure how to participate in this application of the Semantic Web, but according to Arlitsch (2014b), ensuring a library's institutional identity appears on a Google Knowledge Card can be as simple as creating a Wikipedia entry. Wikipedia, through the intermediary of DBpedia, is one of Google's go-to sources of trusted data for its Knowledge Graph (Arlitsch, 2014b). DBpedia is a linked data initiative that transforms Wikipedia's vast volumes of crowd-sourced information into a knowledge base of structured data (DBpedia, 2015). This knowledge base is then used by Google, among other sources, to populate search results with Knowledge Cards. By creating a detailed and accurate entry in Wikipedia, librarians can define the semantic identity of their institutions, and gain control of the way their libraries' information is displayed in Google searches (Arlitsch, 2014b). Similarly, claiming ownership of the library's entry in Google Places or Google My Business can help librarians control their libraries' identity online, and ensure that Google's Knowledge Graph displays correct information. Of course, these techniques are limited to the overall identity of the library, providing general information such as branch locations and hours. An accurate Wikipedia entry will not have an impact on the visibility of individual library resources. Nevertheless, having a well-defined semantic identity online,

and appearing in Knowledge Graph and Satori, are important first steps in participating in the Semantic Web.

Linked data initiatives are springing up worldwide, as shown in the OCLC's recent survey of linked data implementers (Smith-Yoshimura, 2014). However, the timeline for broad application of linked data is unclear. In the meantime, libraries are faced with an immediate visibility issue. This results in a split focus between next-generation linked data and the current realities of OPAC software.

2.3 Gaps in the Literature

The literature on libraries' search engine visibility leaves an important question unanswered. If being visible on search engines has such a powerful impact, why are libraries not rushing to make it happen? Onaifo and Rasmussen showed a majority of Ontario libraries with quite low numbers of indexed pages (2013). In her admittedly cursory examination of the /robots.txt files of Canadian public libraries, Blandford (2015) found that approximately half of the libraries she investigated had /robots.txt files that blocked search engines from indexing the majority of library resources (p. 10) (see Section 1.2.2.1 for an explanation of /robots.txt files). My own examination of participating libraries' /robots.txt files (see Section 4.5 and Appendix H) shows that three of the ten libraries participating in this study have /robots.txt files that block search engine crawlers entirely. What stands between these resources and search engines?

Arlitsch offers one possible answer to this question, in noting libraries' attachment to forms of metadata which are not compatible with search engines (Arlitsch, 2014a), and identifying a lack of SEO awareness among library staff (Arlitsch, O'Brien & Rossman, 2013). Thurow's (2015) identification of negative perceptions of search engine optimization also sheds some light on this question. However, these are personal observations, rather than empirical evidence. Blandford (2015) identifies negative attitudes about search engines themselves as a barrier, based on LIS literature, but the collection of original data was outside the scope of her article. If Arlitsch, Thurow,

Blandford, Breeding, and their colleagues are correct in stating the importance of SEO for libraries, then identifying the reasons behind libraries' continued existence in the deep web is essential. What practices, attitudes, or other obstacles are stopping libraries from reaching out to patrons using search engines? The first step in dispelling these barriers is identifying them.

There is a similar lack of research surrounding factors which could potentially enable libraries to be more visible. Breeding (2014) and Fichter and Wisniewski (2014) outline practical steps to enable visibility at a technical level, which is an important and helpful step. However, these papers do not address the need for higher-level, systemic enablers at an organizational or even profession-wide level. Arlitsch, O'Brien, and Rossmann (2013) identify increased integration into organizational goals as one potential remedy for libraries' current lack of visibility, and Thurow (2015) suggests educating and informing librarians in order to combat negative perceptions. Linked data is suggested as a potential technological enabler (see Section 2.2.4), but this potential is still largely theoretical. These suggestions provide a good starting point for further research, but none of the authors mentioned above claims to base their recommendations on empirical research.

Another significant gap in this body of research is caused by a narrow focus on digital collections. Often, developing a library's digital presence is seen as a movement away from local service towards more global, network-level service. For example, in a paper that assesses the present and predicting the future of collections development, Dempsey, Malpas and Lavoie (2014) highlight the distinction between "outside-in resources, where the library is buying or licensing materials from external providers and making them accessible to a local audience […], and inside-out resources that may be unique to an institution (for example, digitized images or research materials) where the audience is both local and external" (p. 410). This dichotomy has led libraries' SEO initiatives to be focused on digital materials, rather than physical collections. Indeed, almost all of the literature referenced above is about SEO for digital collections.

This focus misses the fact that search engines and online visibility in general can be important *local* outreach tools as well. Of course, digital resources depend more heavily on online discovery, because they exist entirely online. Nevertheless, any pizzeria owner can attest to the importance of web presence in connecting local customers with physical products. As Blandford (2015) notes, standing on an unfamiliar street corner, one can Google "pizza" on a cellphone and find the nearest local vendors. Similarly, one should be able to search for a book and find the nearest copy, which is, in all likelihood, at the local library (Blandford, 2015). While it seems likely that this visibility would have a strong positive effect on the use of libraries' physical collections, there has been no research as yet to support or disprove that hypothesis.

This review found increasing interest in the topic of SEO for libraries, and several studies on the potential impact of SEO on the findability of collections (as outlined in the above literature review), but no original research investigating the reasons for libraries' slow to non-existent adoption of SEO as an outreach practice. In addition, the literature to date focuses almost entirely on academic libraries and digital repositories, with very little written on public libraries or physical resources. Blandford (2015) is the only author to address Canadian public libraries specifically; her article provides a convincing call to action, and points to the need for further research, but again stops short of collecting original data. This study is the first to address search engine visibility for Canadian public libraries through the collection of qualitative and quantitative data.

2.4 Summary of Literature Review

This chapter has described different approaches to search engines in the LIS field, from the perception of search engines as a threat, to the idea of search engines as a critical outreach tool. The Semantic Web and linked data have been introduced as one possible approach to making library resources more visible in SERP, while more traditional SEO techniques are also mentioned. There is increasing interest in the potential of search engines to make library resources more visible to the public; however, little research has been performed to investigate the current visibility of Canadian public library resources, and no research has been done to investigate the factors that influence the visibility of

those resources. This study hopes to begin filling that gap in the literature by shedding light on Canadian public libraries' interactions with search engines.

CHAPTER 3 – METHODOLOGY

This section explains the research methods employed to gather data about the three research questions stated in Section 1.1. Section 3.1 provides an overview of the study design, and the reasons this design was chosen. Section 3.2 describes the recruitment of participants. Section 3.3 describes the research instruments used to collect a robust combination of qualitative and quantitative data. In Section 3.4, the techniques used to analyze each type of data are described. Section 3.5 addresses the research ethics considerations involved in undertaking this project. Section 3.6 outlines the efforts made to ensure validity and reliability, and discusses the study's limitations.

3.1 Study Design

Because this is a relatively new area of study within the LIS field, there were many possible approaches to take in collecting data. For instance, case studies could be performed to test the viability of implementing SEO solutions in real-life library situations; surveys could be administered to test librarians' overall sentiments toward search engines and SEO; quantitative studies could be designed to test searchers' potential reactions to seeing more library resources in SERP. However, because my research questions involve so many unknowns, and there are so few pre-existing concrete data points, I decided that the first step in illuminating this issue was to investigate barriers, motivators, and enabling factors through qualitative interviews with librarians. Without actually discussing the topic with librarians, and discovering the factors they feel are impacting their libraries' visibility, research on this topic risks trying to answer the wrong questions, or providing solutions to the wrong problems.

A mixed-method research design was employed in order to provide a holistic picture of the landscape surrounding search engine visibility in Canadian public libraries. Collecting qualitative data allowed me to explore themes as they arose in interviews, uncovering factors which may be motivators, barriers, or enablers for search engine

visibility in Canadian public libraries. Collecting quantitative data allowed me to triangulate and strengthen the validity of the themes that arose in my qualitative analysis. Qualitative data was collected through a series of semi-structured interviews with library staff members across Canada. Quantitative data was collected in the form of site-specific searches performed for each library's OPAC domain, as well as website usage statistics shared by each library. This approach allowed for the presentation of a well-rounded picture of each library's interaction with search engines. Triangulation, using quantitative data points to balance qualitative assessments, allowed for a richer and more robust analysis of the data (Östlund, Kidd, Wengström & Rowa-Dewar, 2011).

With little existing research to guide my methodology, I decided to perform a pilot study (Interviews 1 and 2), in order to test my research instruments and the usefulness of the data collected. These interviews, along with accompanying quantitative data, were successful in collecting rich data, and in allowing me to hone my research collection instruments (see Dickinson & Smit, 2015). After confirming the viability of my methodology, I expanded my study to include eight more libraries, and conducted ten additional interviews with library staff. I also collected accompanying quantitative data. This approach allowed me to ensure that my research instruments were collecting useful data, and to perform some initial analysis to help guide further data collection and analysis.

3.2 Recruitment of Participants

Ten public libraries across Canada participated in this study. In order to obtain a wide range of data, and keep the data as comparable as possible, I limited my data sample to library systems with a minimum of ten branches, and included as many Canadian geographical regions as possible. Interview participants were chosen in consultation with the library in question, based on candidates' expertise and roles within each organization. My choice among library systems was also determined in part by the availability and willingness of appropriate participants within each system. Of the libraries contacted, two declined to participate in the study, while ten agreed to participate. The participating

libraries will not be identified, in order to protect participant confidentiality (see Section 3.5.2). However, a general overview of the libraries involved is provided in Table 2 (Section 4.1.1), and an overview of participants' areas of expertise is given in Table 3 (Section 4.1.2).

Because resource visibility in search engine results is a relatively new consideration for libraries, not all library systems have incorporated responsibility for this issue into their organizational structures. Therefore, it can be difficult to identify individuals within an organization who are knowledgeable about the subject. For this reason, I began my search for participants by contacting the senior management of each library system (most often the CEO), and requesting their help in identifying an appropriate participant within their organization, based on expertise and job description. Contacting senior management ensured that I had permission from senior management to include each library's data in my study.

Three recruitment e-mails were used, depending on the situation. A generalized recruitment e-mail (see Appendix C) was sent to senior management within each library, which could be circulated among library staff until an appropriate participant was identified. Once an appropriate participant had been identified, I was either put in contact with the potential participant by the library's management, or approached the potential participant directly using a personalized recruitment e-mail (See Appendix D). In the case of Libraries 1 and 2, I approached each library a second time to request a follow-up interview. This was done using a personalized e-mail (see Appendix E). However, in both cases, the original participant was either no longer part of the institution, or had entered a different role and was no longer the most appropriate person for the interview. This made it necessary to seek out new participants for each library, using the procedures outlined above.

3.3 Research Instruments

3.3.1 Interviews with Library Staff

Qualitative data was collected in the form of semi-structured interviews with library staff, from major public library systems across Canada, who have expertise regarding the discoverability of library collections through search engines (see Sections 4.1 through 4.4 for results). See Appendix A for a list of interview questions. Because the interviews were semi-structured, this list of questions was used as a guide; not every participant was asked every question on the list, and additional questions were added in the course of conversation in order to follow up on themes as they arose. Thirteen participants from 10 libraries across Canada were interviewed.

In general, one participant was interviewed for each library in the study. However, in the case of Libraries 1 and 2, one participant was interviewed from each library in early 2015, and a second participant was interviewed from each library in early 2016, which helped provide a sense of each library's development on this issue over time. Please see Appendix B for a list of follow-up interview questions. The initial interviews with Libraries 1 and 2 also acted as a pilot study, in which the study's methodology was tested and refined. Additionally, in the case of Library 8, two participants wished to take part in the same interview, and thus both took part in the same conversation.

Interview participants will be identified by a number that corresponds to their library. So, for instance, Participant 7 (P7) was a staff member at Library 7 (L7) at the time of the interview. In cases where libraries are associated with multiple participants, participants are differentiated as follows: P1a, P1b, P2a, and so on. This avoids confusion, maintains the link between participants and their libraries, and protects the privacy of both participant and organization.

3.3.2 Site-Specific Searches

Site-specific searches were performed in Google in order to collect a combination of qualitative and quantitative data (see Section 4.5 for results). For each library, the

command “site:” followed by the library’s OPAC domain name (e.g., “site:www.catalogue.ca”) was entered into the Google search box. This method was chosen, as opposed to a general keyword search for each library in Google, because it allowed me to differentiate between libraries’ OPAC domains and their main websites. Many of the libraries in this study had high search engine visibility for their main websites, but not for the resources displayed in their OPACs, making it important to differentiate between the two different domains. A general keyword search for the library in Google would retrieve results from libraries’ main websites, their OPAC domains, and other, unrelated sources online. Due to the large volume of results retrieved in any Google search, such a general search would only allow me to examine the library results that rose highest in SERP. Due to the generally speaking low visibility of libraries’ OPAC domains, I would be unlikely to encounter many OPAC pages in the top search results. While a general keyword search might be a good way to assess which, if any, library pages had succeeded in rising high enough to be visible in SERP, it would not provide a complete view of how many OPAC pages were being indexed, regardless of whether they had risen to the top of search results. A site-specific search, on the other hand, allowed me to view the results for every page Google had indexed for each domain at the time of the search. It also ensured that my results were more comparable between libraries, avoiding any bias that might be introduced if Google personalized results based on my location or search history.

By performing these site-specific searches of libraries’ OPAC domains in Google, I collected data on how many of the information resources contained in each library’s OPAC were indexed by Google at the time of the search. While not exact, these searches provide a rough estimate of each library’s visibility through a quantitative assessment of the number of pages indexed, and a qualitative assessment of the quality of indexed results. Because the results returned by site-specific searches appear in the same format as they would in any other SERP, they allow an assessment of the usefulness of results from an online searcher’s perspective. While knowledge of the number of pages indexed does not amount to a complete assessment of search engine visibility, it does offer a snapshot of how each library’s resources might appear to a searcher. Site-specific

searches were conducted within a few weeks of each interview, in order to ensure that each library's results would provide a useful counterpoint to the interviews (please see Appendix I for a table summarizing the time period of each search). In the case of libraries 1 and 2, where two interviews were conducted at different times, a separate site-specific search was performed at the time of each interview. L2's Libhub site was only included in the second set of searches, because at the time of the original interview with L2 their Libhub site was not yet up and running. For the purposes of comparing between libraries, the most recent results for L1 and L2 were used.

3.3.3 Website Traffic Data

The second type of quantitative data collected consists of website usage statistics provided by each library (see Section 4.6 for results). When possible, data was obtained about the source of hits and source of referrals to each library's main site, as well as each library's OPAC domain. This data allowed me to investigate patterns of usage for libraries' main websites and OPAC domains, and assess how users were reaching each library's resources. However, each participant in this study collects different amounts of website traffic data in different ways, depending on the amount of data available to them and the tools to which they had access for each domain. In addition, depending on each library's policies concerning data and privacy, participants could share differing amounts and types of website traffic data with me, for differing time periods and in differing levels of detail. In all cases, I attempted to collect data from a period near to the date of the interview, however the actual period covered by each data set was decided by participants. This made it difficult to directly compare libraries usage statistics, and in some cases also weakened comparisons between participant interviews and website traffic data. However, some comparison is possible between some of the libraries. The website traffic data collected also allows for individual assessments of current website traffic patterns surrounding commercial search engines for each library, which provides a point of comparison during the analysis of the qualitative data.

3.4 DATA ANALYSIS

3.4.1 Analysis of the Interviews

Interviews with library staff were analyzed using thematic analysis techniques, in order to uncover themes, and discover the relationships between themes (Bazeley, 2013; Ryan & Bernard, 2003). A tentative set of broad initial themes was gleaned from the above literature review. Based on various authors' speculations about why library resources remain less visible in SERP (see Section 2.3), I began analysis searching for potential barriers in the form of attitudes, organizational issues, and technical difficulties. The current discourse on the topic also prompted me to watch for themes involving librarians' perceptions of search engines, in order to confirm or deny speculation about negative perceptions of search engines as a potential barrier. Similarly, my analysis of potential enablers was initially informed by the above literature review. My initial analysis of interviews 1 and 2, which were performed as a pilot study in 2015, also provided a tentative framework for codes pertaining to barriers to visibility and general perceptions of search engines as I moved forward with the rest of the interviews.

After anonymization, transcripts were uploaded into the cloud-based qualitative analysis software Dedoose (http://www.dedoose.com/), which allows for the creation and manipulation of interconnected codes. Through close readings of each transcript, additional codes were developed to describe concepts that recurred throughout multiple interviews. This process resulted in approximately 75 codes. Similar codes were amalgamated, and these codes were then categorized into larger thematic groups in order to understand the relationship that each concept had to my research questions. During this process, three overarching themes developed in direct relation to the three research questions outlined in Section 1.1: Motivators of visibility (see Section 4.2), barriers to visibility (see Section 4.3), and enablers of visibility (see Section 4.4).

Table 1 shows an overview of the codes that emerged during analysis. This table outlines the general categories with which these codes were associated, the number of times each code was applied, and the section(s) of the Results chapter where each code is discussed, if applicable. Certain categories involved codes which connected to various different

themes. For instance, the "Perceptions of search engines" category involved negative perceptions, which were relevant to attitudes as a barrier to visibility, as well as positive perceptions, which were relevant to attitudes as an enabler of visibility. An examination of this table also reveals that some codes were used very frequently, while others occurred less frequently. This is because each participating library is at a different stage in the development of search engine visibility for their library resources. As discussed in Section 3.6.4, some factors are perceived by many participants, while others are perceived by only a few participants, because a library in the midst of actively pursuing visibility will encounter different factors than a library that has not yet considered search engine visibility as a possibility. Due to the relative lack of research in this area to date, I believe that codes with a relatively low number of occurrences are worth discussing in order to begin exploring this complex issue. Reporting the number of occurrences for each code in Table 1 allows the reader to note codes with a higher or lower number of occurrences, and take that information into account when assessing the results described in Chapter 4.

Table 1. Overview of Codes

Theme	**Sub-category**	**Code**	**Number of Occurrences**	**Section of Results Chapter**
Drivers/ motivators		Important on principle	1	4.2 Motivators of Visibility
		Libraries need to adapt	12	4.2.1 Adapting to Changing Search Behaviour
		Public awareness of deep web, libraries not in search results	2	
		Changing search behaviour	11	
		Bring more traffic to library websites	19	4.2.2 Reaching Non-Library-Users
		Importance to overall image of library as org	8	

Theme	Sub-category	Code	Number of Occurrences	Section of Results Chapter
		Libraries can contribute to search engines	10	4.2.3 Contributing to the Quality of Information on Search Engines
Barriers	Attitudes	Distaste for commercial-ness	4	4.3.1.1 Search Engines as Commercial and Untrustworthy
		Sense that it's not possible	14	4.3.1.2 Visibility as an Impossible Goal
		Unknown consequences	8	4.3.1.3 Unknown Consequences
		Don't need the exposure	10	4.3.1.4 Lack of Motivation to Become More Visible
		Need to protect data/privacy	13	4.3.1.5 Concerns About Data Privacy and Security
		Investment in previous practices/formats	5	4.3.1.6 Investment in Current Practices and Formats
	Organizational issues	Not integrated into org. goals	15	4.3.2.1 Lack of Integration Into Organizational Goals
		Lack of comfort/understanding/awareness of tech	17	4.3.2.2 Lack of Comfort With Technology Throughout Organizations

Theme	Sub-category	Code	Number of Occurrences	Section of Results Chapter
		Lack of buy-in throughout org	7	4.3.2.3 Lack of Buy-In Throughout Organizations
		Multi-departmental issue	10	4.3.2.4 Lack of Integration into Departmental Structures
		It's nobody's job	4	
		Lack of policy	23	4.3.2.5 Lack of Policy
	Technical difficulties	Current catalogue can't handle it	11	4.3.3.1 Difficulties Being Indexed
		Data in silos	5	
		rapidly changing collection	4	
		Lack of links to catalogue content	2	
		Faceted search	4	4.3.3.2 Difficulties Rising to the Top of SERPs
		Need to be in the top results	12	
		Geolocation or lack thereof	4	
		MARC records	9	4.3.3.3 General Technical Difficulties
	Limited resources (staff, time, money)		37	4.3.4 Limited Resources
	Vendors (as barriers)		65	4.3.5 Vendors
Data (General discussion of data, as both barrier and enabler)			43	4.3.1.5 Concerns About Data Privacy and Security; 4.4.4 Technological Enablers

Theme	Sub-category	Code	Number of Occurrences	Section of Results Chapter
Enablers	Attitudes	Libraries cooperating	13	4.4.1 Inter-library Cooperation
		SEO as goal/desireable	30	4.4.2 Attitudes
	Organizational	Culture of innovation in organization	2	4.4.3 Organizational Enablers
		Flexibility/ willingness to take risks in organization	4	
		High priority in organization	1	
	Technological	Access to accurate data/analytics	10	4.4.4 Technological Enablers
		Catalogue search integrated into main site(s)	5	
		Clean data to work with	2	
		Control over OPAC/related data	2	
		Up to date website, ILS	2	
		use of social media	9	
	Vendors(as enablers)		29	4.4.5 Vendors
General points of comparison		Amount/quality of resources being indexed	16	
		How users find library resources(/OPAC)	22	
		Job Description	29	4.1.2 Participant Profiles

Theme	Sub-category	Code	Number of Occurrences	Section of Results Chapter
		Level of focus on issue in organization	34	4.3.2.1 Lack of Integration into Organizational Goals; 4.4.3 Organizational Enablers
		Impact of being indexed/visible	19	4.2.2 Reaching Non-Library-Users
Perceptions of search engines		Change rapidly	14	4.3.1.1 Search Engines as Commercial and Untrustworthy; 4.3.2.5 Lack of Policy
		Commercial, motivated by profit	19	4.3.1.1 Search Engines as Commercial and Untrustworthy
		Threat to libraries	18	
		Don't care about libraries	4	
		Untrustworthy	10	
		Different from libraries	13	
		All-seeing, collect data on everyone	4	4.3.1.5 Concerns About Data Privacy and Security
		Control info access	11	4.2.1 Adapting to Changing Search Behaviour
		Popularity/ubiquity	19	
		Shape user expectations	12	
		Generational perception of search engines	3	4.3.2.2 Lack of Comfort With Technology Throughout Organizations

Theme	Sub-category	Code	Number of Occurrences	Section of Results Chapter
		Hard to control (what's indexed etc)	10	4.3.3.3 General Technological Difficulties
		Models of usability, easy to use	12	4.4.2 Attitudes
		Useful tools	12	
		Not as good quality as library tools, poor quality info	10	4.2.3 Contributing to the Quality of Information on Search Engines
Semantic Web		Semantic Web (generally)	28	4.2.3 Contributing to the Quality of Information on Search Engines; 4.4.5 Vendors
		LibHub	49	
		Google Places	14	4.3.3.3 General Technological Difficulties

3.4.2 Analysis of the Site-Specific Searches

The data gathered using site-specific searches was used to assess each library's visibility in Google at the time of the interview. In addition to determining the total number of pages indexed for each library, I assessed the quality of each library's search engine results, using information provided by Google about SEO best practices (Google, 2010) as a guide. The factors included in this quality assessment were:

1. Whether the /robots.txt file allowed crawlers access to content.
2. The amount of detail included in each result's rich snippet, and the appropriateness of that information in helping a searcher evaluate results.

3. The amount of detail included in each result's title, and the appropriateness of that information in helping a searcher evaluate results. For example, did titles include the library's name, or only the name of the discovery layer or OPAC vendor?
4. The different types of resources included in results: Did results include a diverse range of the resources offered by the library?

The number of pages indexed, as well as the quality of the results, were used to categorize libraries into five broad categories on a scale of overall visibility, from no results whatsoever to very detailed and complete SERP (See Section 4.5). This information was then used as a triangulation point for participant interviews, providing a point of comparison with participants' perceptions regarding their libraries' visibility.

3.4.3 Analysis of the Website Traffic Data

The website traffic data provided by each library was mainly used as an added point of comparison, along with the site-specific searches, in order to better understand the information provided by interview participants. Data on the source of hits to libraries' different domains allowed an assessment of the path users take to reach library resources. This assessment was then used as a point of comparison to help confirm the accuracy of interview results. While a conclusive comparison among libraries based on website traffic is impossible, due to the differing types and amount of data provided by each library, a high-level comparison of website traffic data did reveal some interesting patterns which suggest avenues for future research.

3.5 Research Ethics

The methodology employed in this study was developed in accordance with Tri-Council Policy Statement *Ethical Conduct for Research Involving Humans* guidelines for ethical research (CIHR, NSERC & SSHRC, 2014), and approved by Dalhousie University's Research Ethics Board. The following subsections outline the procedures used to ensure informed consent (Section 3.5.1), confidentiality and anonymity (Section 3.5.2), and risk mitigation (Section 3.5.3).

3.5.1 Informed Consent

Informed consent was ensured by providing an electronic copy of the consent form (See Appendix F) during the recruitment process, in advance of the scheduled interview. As phone interviews were the primary data collection instrument, requiring participants to print, sign, scan, and email a consent form in advance would have added an additional burden of time to participants. Therefore, rather than requesting a physical signature, I obtained oral consent from participants at the beginning of each interview. In order to ensure that the consent form had been fully read and understood, I summarized paragraphs 2-8 of the consent form orally before beginning each interview, and gave participants an opportunity to voice any questions or concerns, or to withdraw from the study, before asking for their consent. Their oral consent was recorded and remained with the data collected until the recording was transcribed and the transcription anonymized.

3.5.2 Confidentiality and Anonymity

Interview questions related to participants' professional lives, and as such may include personally identifying information. Interview data was digitally recorded, with participants' permission, and transcribed. Interview recordings were stored locally on a secure, password protected laptop. Transcriptions were only linked with the original recordings and quantitative data through assigned identification numbers. Transcribed interviews were kept in a separate location from the recorded interviews and the quantitative data provided by each library, in order to preserve the confidentiality of participants. No paper copies of the transcripts, or handwritten notes, were made.

It was not possible for interview participants to be completely anonymous during interviews; however, all personally identifying information, including identification of workplaces and geographical locations, was redacted during transcription. By keeping not only the participants' identities, but also their workplace and geographical location confidential, I minimized any risk that participants could be identified based on workplace or location and job description.

The quantitative data provided by each library may contain confidential information relating to each institution. In order to protect this information, it was stored securely on a password protected laptop, and only anonymized information is reported. In addition, this usage data often contained information, such as domain names, which could identify the institution in question. As such, linking this data with interviews could risk identifying the interview participant through their institution or location. To mitigate this risk, the data provided by libraries is only published in anonymized form. For example, where the raw data contains specific domain names which identify an institution, when presenting an analysis of this data, domain names are replaced by generalized identifiers such as "Library 3's Main Site". The raw data provided by each library was stored separately from the anonymized interview transcripts; these two datasets were only linked using assigned identification numbers.

All practical precautions were taken to ensure that the findings of this study are not linked in any way to participants or their workplaces, and that the combined analysis of quantitative data and interview responses does not risk identifying participants or their institutions.

3.5.3 Risk Assessment

The questions asked in the interviews relate directly to each participant's job, as well as the policies and practices employed in their workplace. Because of this, participation might pose some risk to participants' professional and social standing. For example, a participant might express an opinion that agrees or disagrees with a policy of their organization.

In addition, sharing quantitative data in the form of website usage statistics may involve some risk for participants' libraries, by allowing me access to potentially confidential institutional information. Some organizations may be uncomfortable allowing an individual from outside their organization to view and analyze data without their input or

context, or may be concerned that once outside of their control the data will not be properly safeguarded.

The estimated probability of these risks is low, and they do not exceed risks encountered in the regular activities of participants. Participation in this study was voluntary, and participants were given the opportunity to withdraw at any point if they felt uncomfortable.

3.6 Validity, Reliability, and Limitations

3.6.1 Validity

Validity was ensured in several ways throughout this study. First, I have identified my own assumptions and biases in order to acknowledge the possible impact they may have on my analysis of the data (see Section 3.6.2.1). Because I have strong opinions about the importance of search engine visibility for libraries, this "researcher reflexivity" is important in allowing readers to understand my perspective and how it may have shaped my research (Cresswell & Miller, 2000, p. 127). Peer debriefing has also been an important aid to validity. During the course of this project, I have discussed my methodology, ethical concerns, and analysis techniques with a mentor on a weekly basis. This allowed me to challenge any initial assumptions I may have made, and consider the validity of my methods on an ongoing basis. The validity and reliability of the research collection instrument used during interviews was tested in interviews 1 and 2, which were performed as a pilot study in 2015. This allowed me to assess the effectiveness of the study's methodology, and fine-tune the interview guide before continuing with the rest of the interviews.

Validity was also ensured through triangulation. Collecting multiple types of quantitative data during the same period as the interviews provided valuable points of reference to inform my analysis of the qualitative data. This quantitative data helped me understand the current visibility of each library based on concrete data points, which in turn allowed for a better understanding of participants' perceptions about the visibility of their

institutions, and a more holistic understanding of the themes that emerged during analysis.

3.6.2 Threats to Validity

3.6.2.1 Assumptions

This research is undertaken with the underlying assumption that increased visibility for library resources through search engines would be a positive thing for libraries and the community members they serve. As the literature review shows (see Section 2.1.1), this view is not universally held within the LIS community. As a relatively young researcher who has made enthusiastic use of search engines since the age of ten, I may be more likely to see the positive aspects of search engines, and less sensitive to their intrusive or problematic aspects. Quite simply, my comfort level with search engines in general may be much higher than the comfort level of some of the interview participants. Approaching search engine visibility as a positive goal for libraries has shaped both the data collection instruments, and the analysis of the data. This approach provides a useful perspective on the issue of library visibility in search engine results, but should be kept in mind during the consideration of my findings.

3.6.2.2 Limitations

Including ten major library systems from across Canada provides a broad sense of the landscape in Canadian public libraries, but the libraries chosen may not be representative of all Canadian public libraries. This sample-size is still relatively small, and general conclusions about all Canadian libraries cannot be made based on this data. Approaching only library systems with ten or more branches increases the comparability of results, but may reduce the study's ability to speak to the concerns of smaller library systems.

In addition, several aspects of data collection must be noted as causing potential limitations to the validity of this study. First, the fact that interviews were performed over the course of a year caused both the qualitative and quantitative data to be collected at different times for each library. This limits comparability of data; for example, it is

possible that conducting interviews over a longer time span resulted in later participants having higher awareness of the issues being discussed, simply because these issues had had time to become more commonly discussed in the LIS community. This varying time frame was a logistical necessity when arranging interviews with busy individuals in many different institutions, but in an area of study that is changing rapidly, having such a protracted time frame is problematic. However, this limitation does not negate this study's ability to explore and identify perceptions surrounding search engine visibility. In addition, returning to Libraries 1 and 2 at the end of the study and collecting additional data allowed me to assess any potential changes over time that might have occurred in the two institutions that participated at the earliest times.

Second, the website traffic data collected varies from library to library, which makes it difficult to make conclusive comparisons between libraries based on this portion of the data. This is partly because libraries do not all collect the same type or quantity of data; for instance, P10 was unable to share website traffic data for his library's OPAC domain, because he is technologically unable to collect such data. This is also partly because libraries have differing levels of comfort in sharing website traffic data. Issues surrounding patron privacy and the collection of potentially personally identifiable information cause some participants to be quite cautious when collecting and sharing data (see Section 4.3.1.5). I bore these concerns in mind in the collection of website traffic data, and as a result, did not require participating libraries to provide it. Rather, I encouraged each participant to share as much website traffic data as they felt was congruent with their library's policies and their comfort level. This resulted in data that is often not comparable between libraries, and some libraries declined to share any website traffic data at all. However, as the website traffic data was mainly intended to act as a point of comparison with each participant's interview data, rather than as a way to make comparisons between libraries, this limitation was deemed acceptable in order to respect participants' concerns about patron privacy.

Another potential limitation was introduced during data collection for Library 8. Having two people take part in the same interview for Library 8 may have caused the interview to

yield somewhat different data. It is possible that these two participants would have responded differently to questions if they had been alone in the interview. Because both participants work in the same organization, having a co-worker present might have introduced some reticence, or caused participants to feel less comfortable expressing their true views. However, because the two people in question do not work directly together, and are at equal levels of responsibility in the organization (i.e. neither one can be seen to have authority over the other), these risks are minimal. Nevertheless, this limitation should be kept in mind in considering the data collected from Library 8.

Due to the limitations outlined above, this study cannot claim to be a valid representation of all Canadian public libraries. However, the data collected does show that the factors identified are perceived within multiple institutions across Canada. This information provides a much-needed basis for further investigation in an area lacking empirical evidence and original research. Identifying the specific barriers experienced by at least one staff member in various Canadian public libraries paves the way for more focused future research investigating potential solutions to these barriers. Similarly, identifying the factors identified by participants as potential enablers of visibility paves the way for future research examining their effectiveness as enablers, and their relative importance to the success of search engine visibility initiatives.

3.6.3 Reliability

The interviews carried out during this study were digitally recorded and carefully transcribed to ensure that the data reported accurately depicts participants' statements. This is important in ensuring the reliability of qualitative studies (Cresswell, 2007). Ensuring that the interviews were carried out in a consistent fashion also enhanced the reliability of this study. The use of an interview guide (see Appendix A) ensured that although the interviews were semi-structured, allowing me to pursue fruitful lines of questioning depending on participant responses, all participants were asked the same set of standard questions as a baseline.

3.6.4 Threats to Reliability

As an exploratory study with a relatively small sample-size, this study does not claim to reliably capture the views of all Canadian public librarians, or even the views of all the staff members in each participating library. Another member of the same institution might give different answers. Indeed, because this is a rapidly evolving issue, the same participant might give different answers if asked the same questions at a different time.

The results reported are also affected by the fact that each participating library is at a different stage of development when it comes to search engine visibility for library resources. This threatens the study's reliability in terms of the internal consistency of results (Leedy & Ormrod, 2013). Despite consistency in data collection, the same interviews did not yield similar results across all participants. Different participants have different perceptions based on their unique situations; factors identified by one participant may not be perceived by all the other participants. Some librarians have encountered different barriers and enablers than others because they are actively pursuing visibility, while other librarians are not. Making an attempt to be visible naturally uncovers barriers and enablers that might not otherwise be apparent. This means that in some cases, the factors identified recur broadly across most of the participating libraries (for example, limited resources as a barrier in Section 4.3.4), while in other cases factors are mentioned by only a few participants (for example, the organizational enablers in Section 4.4.3). Please see Table 1 in Section 3.4.1 for an overview of codes used and their relative number of occurrences throughout interviews.

Studying this issue in the midst of its rapid evolution, particularly with a relatively small sample size, necessarily involves a certain amount of unevenness in the data collected. All that can be said, without repeating this methodology with more libraries to gain more data, is that the less-commonly-mentioned factors reported here are perceived by at least two or three librarians in two or three institutions. Given that little to no research has yet been performed on this topic, these less frequently occurring themes are still worthy of study, and have the potential to inform further research.

3.7 Summary of Methodology

This chapter has detailed the methodology employed to collect and analyze data in order to answer the research questions stated in Section 1.1. The study's overall design was described, including methods for recruiting participants and measures taken to comply with research ethics requirements. An overview of the research instruments used to collect a combination of qualitative and quantitative data was provided. Data analysis techniques were discussed, and a detailed overview of the codes used during analysis was provided in Table 1 (Section 3.4.1). Methods for ensuring validity and reliability were described, and limitations to the study's validity and reliability were acknowledged. By providing a detailed description of my methodology, this chapter is intended to help the reader make a balanced and critical assessment of this study's findings and their implications.

CHAPTER 4 – RESULTS

In this chapter, the results of participant interviews, site-specific searches, and website traffic data are reported. Section 4.1 provides an overview of participating libraries and interview participants. In Sections 4.2, 4.3, and 4.4, the results of the semi-structured interviews are reported. These results are structured according to three overarching themes, which correspond to the three research questions laid out in Section 1.1: motivators of visibility, barriers to visibility, and enablers of visibility. Structuring the interview results in this way ensures that each research question is addressed thoroughly. Section 4.5 outlines the results of the site-specific searches performed for each library. Based on the results of the site-specific searches, libraries are categorized based on their current level of visibility, and each category is discussed in detail. Section 4.6 reports the results of the website traffic data provided by each library, and these results are compared with participants' perceptions about how patrons currently find their resources. These three different sets of results, taken together, provide a detailed picture of the current state of search engine visibility in the participating libraries.

4.1 PARTICIPANT PROFILES

4.1.1 Library Profiles

Ten public library systems across Canada participated in this study. The sample includes urban systems based in larger Canadian cities, as well as some province-wide systems and consortia. Table 2 provides an overview of participating libraries, including their size, the providers of their OPAC software, and whether they are participants in the Libhub Initiative (see Section 2.2.4 for a discussion of the Libhub Initiative). Libraries have been categorized as small, medium, and large, based on number of branches. Small systems have between 10 and 20 branches, medium systems have 21 to 60 branches, and large systems have 61 to 150 branches. In order to avoid reporting any details that might, when combined with other information, identify participants, this table shows the different types of libraries that participated without linking that data to specific libraries.

Table 2: Overview of Participating Libraries

Branches	Type of system	OPAC/Discovery Layer Vendor	Libhub Adoption
Small	Urban	Aquabrowser Library	No
Small	Urban	BiblioCommons	Yes
Small	Urban	Sirsidynix Enterprise	No
Small	Consortium	Own system	No
Large	Province-wide	BiblioCommons	No
Large	Urban	Own system	No
Medium	Urban	BiblioCommons	No
Medium	Urban	BiblioCommons	Yes
Medium	Urban	BiblioCommons	No
Large	Province-wide	Horizon Information Portal (Sirsidynix)	No

4.1.2 Participant Roles

Because search engine visibility tends to involve several library departments (see Section 4.3.2.4 for further discussion of this issue), participants in this study had various different roles. Near the beginning of each interview, participants were asked to provide their job titles, and a description of what their job entails. Because specific job titles, when combined with other data, might risk identifying participants, general areas of expertise are reported here based on participants' descriptions of their jobs. See Table 3 (below) for a summary participants' areas of expertise.

Table 3: Overview of Participants' Areas of Expertise

Participant	Area of Expertise
1a	Web development
1b	Web development
2a	Digital display of library resources
2b	Digital display of library resources
3	Management of library technical systems and collections

Participant	Area of Expertise
4	Administration of library technology services consortium
5	Management of library technical systems (formerly management of library collections)
6	Web development and user experience assessment
7	Management of library technical systems
8a	Management of library collections
8b	Management of library technical systems
9	Management of library technical systems
10	Management of library technical systems

As shown in Table 3, the majority of participants have expertise related in some way to information systems or information technology, while several have expertise relating to library collections. In the case of Library 2, the participants' role in the organization was created specifically to address the online discoverability of library services. Participants also come from a variety of levels within the organization; some work as part of a team to accomplish the library's goals, while others fill roles in various levels of management and administration.

In some cases, participants identify the question of search engine visibility as something they deal with directly as part of their jobs. In other cases, participants state that the issue of search engine visibility does not fall under any position's job description within the organization. For example, the search engine visibility for library resources is an integral part of P2a and P2b's job descriptions. P10's role, on the other hand, includes search engine optimization of the library's main website, but has little involvement in the visibility of the library's OPAC or the information resources it contains. Generally speaking, participants who did not include search engine visibility for library resources in their explicit job description confirmed that their library currently lacks any specific position that addresses the issue.

4.2 Motivators of Visibility

The first step in making library resources more visible on search engines is for libraries to make visibility a goal. Without that drive to become more visible, this aspect of

organizational outreach may never rise high enough on the list of priorities to be addressed at all. As noted in Section 4.3.2, this issue is not always on the radar for libraries. The utility of SEO may seem self-evident. Indeed, P6 feels that regardless of any other considerations,

> It's important just on principle for libraries to be search engine friendly and indexable and discoverable. You never know how that might work out in the long term, and how that might enable different people to develop different specialized search tools or aps, or whatever that might be useful. I mean, it's just the right thing to do to ensure that your data is discoverable to machines. That's just good practice.

However, more pressing motivation than general good practice may be needed to encourage busy organizations with many competing priorities to spend effort and money on search engine visibility. What factors could motivate libraries to prioritize SEO for library resources? Participants in this study mention three factors that are causing search engine visibility to take on increasing urgency: the need to adapt to changing search behaviour (Section 4.2.1), the need to reach out to non-library-users (Section 4.2.2), and the desire to contribute to the quality of information available through search engines (Section 4.2.3).

4.2.1 Adapting to Changing Search Behaviour

As noted in Chapter 1, several studies have shown that search engines have become by far the most popular online information seeking tools. People who might at one time have asked their local librarian to look up a fact are now getting answers from Google within seconds. In the course of these interviews, I found that the participants of this study are very aware of this trend:

> Year after year, we're seeing fewer and fewer people doing anything but going to search engines. At this point, I think it's pretty widely acknowledged that search engines are the place that people are performing their information seeking behaviour. (P2a)

> It's the common theme that people use search engines to find their information. So getting listed, and listed at least on the first page or in the top five of a search engine is very important. (P10)

P10 describes search engines as being so convenient that they sometimes supersede browser bookmarks in retrieving known sources:

> No one really even uses bookmarks anymore. People just search. They may have a favourite – I'm guilty of that myself. I have a bookmark, I have all kinds of bookmarks for different technology things. But instead of just going back to look for what I need in my bookmark, I'll just Google it and pull it up again. (P10)

This awareness of changing search behaviour translates into a growing feeling that libraries must adapt in order to remain relevant. As P4 puts it, "That's where our customers are, and we're not invading that space as much as we need to." Not adapting to the new reality of search is seen as a serious threat to libraries:

> In the long term I think we're at that time where we have to do something about the way people are searching, because they are not thinking about the library the way we would like them to think about the library. (P8a)

When asked if search engines posed a threat to library services, P4 answered:

> I think the biggest threat is libraries not adapting to technology, to ways of life that are changing so rapidly over the last ten to fifteen years, and it's not going to slow down. I think that's the threat. (P4)

The threat posed by failing to adapt is seen as having very real consequences if left unaddressed:

> If libraries don't continue to evolve and grow and look at what is actual information behaviour, and what people want, what do they consider useful, I think we won't continue to exist in the future. (P3)

This is particularly the case because many members of the public may not be aware that library resources aren't present in search engine results, and therefore may assume that

the library does not have the resource in question when it does not show up in SERP. P4 sees this as a serious potential threat. When asked whether he thinks his library's patrons are aware that the library's resources are not indexed by search engines, he states:

> I don't think they think about it. I just don't think that they, I mean this is just a complete guess here, but I'm trying to think about myself as a patron, trying to think about not having the background of libraries, not working in the field, and you know, if I was looking for something, I'm going to Google, because that's where everybody goes, and I find what I need and then I leave. I don't think I'm thinking about the fact that there's more, and it's my public library, and that content's not here. That's the danger. That's why our not doing anything is the threat. (P4)

P8b echoes this sentiment, when asked how many library patrons she thought might be unaware that library resources are not visible in a Google search:

> I would suspect a significant portion because I think even when we've been having the conversations around BIBFRAME internally with staff, a lot our frontline staff don't even necessarily have that awareness, that MARC data is not visible to the web.

4.2.2 Reaching Non-Library-Users

A big part of the adaptive process for libraries involves reaching out to community members who do not currently use library services. As P1b puts it, "If you want to reach new people, especially people that aren't already using the library system, you definitely have to be thinking about this stuff, right?" Speaking of her library's involvement in the Libhub Initiative, P8a stated, "I think the best case scenario is, 'Oh wow, the library has it! Never thought of going to the library.'" P4 said that his library is building its capability to take the necessary steps to "be relevant to somebody who's 15 who's never used the library, and never even thought of it. We want to catch their attention, and we plan to, and we're setting a stage for that."

Many participants were hopeful that if library resources could be visible enough in SERP, they would reach community members who had not previously considered the library as a resource:

> Through serendipity in a search engine result, ideally, somebody who wouldn't know about the library, maybe if your stuff pops up pretty high in a search results list, it'll make them think, "Oh, I'll go to the library! I never thought of that". (P9)

> The big benefit to me of having our catalogue contents indexed by search engines would be if somebody looked up something popular and on their first page they say, "Oh, I'm getting a result from my local library, what's that about? They have it, oh my god, that's great, I don't have to buy it!" (P3)

This need to reach out and be accessible to every member of the community is deeply ingrained in the values of public librarians. Being accessible is not simply about self-preservation; it is also about the core raison-d'être of public libraries:

> As a public librarian, […] you want to be where people are and you want to serve them as best as possible. And so I think search engines are definitely a way to help serve the people as best as possible. (P2a)

4.2.3 Contributing to the Quality of Information on Search Engines

Another core value of librarianship, the desire to ensure people are getting the highest possible quality of information, provides an added motivation for some participants to make their libraries' resources more widely accessible. Along with a nigh-universal acknowledgement of Google's ascendancy as an information seeking tool, participants expressed a widespread sense that information gained through search engines can be unreliable. As P10 states, "There's information, and then there's quality information." This is an area where librarians feel qualified to help:

> One of the older values of public libraries was always, all the information and detailed information, and a real ability to drill down and research a subject in detail. (P7)

Libraries are well-established as a source of quality information and information literacy education. A number of participants in this study feel that libraries can now leverage the high quality information at their disposal to actively improve the quality of information available on search engines. As library data begins to open up to search engines, some participants expect that "libraries can provide online resources that can help search engines answer questions" (P5).

Participants from both libraries taking part in the Libhub Initiative cite a desire to improve the search engines' ability to provide quality of information as one of their motivators in becoming compatible with search engines:

> I think it's really our time, because there's so much traffic on the web, and trying to find, I would say in quotation marks "authoritative sources", it's nice when they could point back to a library. I think it's time that libraries become more engaged. (P8a)

> We have this beautiful data. How wonderful would it be if libraries could share that with search engines? Because there's a lot of inconsistent or sub-par data online right now. (P2a)

4.3 Barriers to Visibility

Based on participant responses, this study identifies a complex combination of barriers to visibility. These barriers have been placed into five broad categories, discussed in the sections below: attitudes, organizational issues, technical difficulties, limited resources, and vendors. Some of the barriers reported here were consciously identified as barriers by participants, while others have been identified by the researcher through a rigorous analysis of the interviews (see Section 3.4.1 for a discussion of interview analysis).

4.3.1 Attitudes

Librarians' attitudes about search engines and SEO initiatives are an important factor in deciding whether an organization will even consider undertaking such initiatives. The participants of this study expressed many opinions, both positive and negative, about

search engines and the potential of SEO to benefit their libraries. Attitudes that might constitute barriers to increased visibility are reported here, and attitudes that might enable search engine visibility are discussed in Section 4.4.2.

4.3.1.1 Search Engines as Commercial and Untrustworthy

One set of attitudes that may have a profound effect on libraries' ability, or even desire, to become more visible in SERP, are negative perceptions of search engines. While most participants express positive feelings about search engines as useful and popular tools, many also express negative views about search engines' commercial focus:

> My opinion of search engines is that they're for-profit businesses. And so you know, their decision-making processes are always going to be, even if not in a tangible way, the bottom line is what drives them. So I guess my opinion of search engines is that I'm fairly skeptical of large for-profit businesses, you know, that they really have the best interests of the public at heart. (P9)

When asked for his opinions about search engines, money is the first thing that comes to P1b's mind: "Just that they're good at making money!" This commercial motivation is seen as necessary:

> It would be unmanageable for somebody to just altruistically index it. I think we probably are in an era where we need those people who can find a business-motive to do this work. To provide something. But I think we have to always recognize, I think it's really just a matter of recognizing and understanding what their motive is. (P7)

Similarly, P9 notes that "search engine companies are self-interested, as they should be; they have shareholders that they answer to." However, the consequence of this fact is that "they don't answer to the public good" (P9). Search engines' commercial focus can cause feelings of distrust:

> I think that they have a strong profit-motive, and that when they talk about, when they do experimental things and they talk about experimental things and they talk about wanting to do good in the world, that they are not actually… here's a very heavily-loaded word, which is too heavily-loaded: they're not actually lying, but they are putting a very good spin – and a strong

> spin – on what is only part of their interest. And I think their strongest interest is their profit-motive. (P7)

Being motivated by profit is seen not only as untrustworthy, but actively opposed to the interests and values of libraries: "Profit interest will always be, at some level, antithetical to a full and complete and transparent discovery interest" (P7). These negative feelings about the commercial aspects of search engines may be explained, at least in part, by a general distaste for commercial concerns among some participants. Not only is commerciality not part of libraries' mandate; it is explicitly and adamantly excluded: "We're not in the business of facilitating commercial activity in our organization" (P7); "Customer implies a business transaction, a commercial transaction. That's not what we're about, so I've never liked it" (P9). Libraries are seen as unbiased and trustworthy institutions that promote open access rather than commercial interests. In this context, commercial concerns are a force to be actively excluded.

When asked what makes search engines different from other for-profit businesses in the LIS world, such as vendors, P9 explains that vendors' success is tied to libraries' success, while search engines have no specific ties to libraries:

> Well, I think in the case of BiblioCommons [a discovery layer vendor], we are the customer of BiblioCommons. They don't exist without public libraries. So in order for BiblioCommons to be successful in their business model, public libraries have to be successful. So we have a shared vested interest in each other's success. Google doesn't have an interest in public libraries' success. They have, there's no vested interest there whatsoever. (P9)

The statements of some participants in this study suggest that librarians feel insignificant to Google: "Before I started library school, I think Google was definitely a pretty positive organization. Now they're, it's shown that they don't really care about libraries that much" (P1a); "If it were just to say, well someday Google will fix a library problem, I doubt that'll... I wouldn't pin my hopes on that one" (P7). This feeling of insignificance, paired with the highly changeable nature of search engines' evolving algorithms, can lead

to distrust: “so many of these Google things are at some level experiments that they could abandon at any time, and […] their profit-motive could change at any time” (P7).

Reactions to the commercial aspects of search engines can be quite visceral: “Every time Google puts an ad of something that I searched maybe two weeks ago it creeps me out a little bit” (P8b). While most participants balance their mistrust of search engines with an acknowledgement of their importance in information discovery, in some cases the perceived commerciality of search engines can be an active barrier to librarians’ willingness to open up library data:

> I’m sort of against giving information to Google, just because Google essentially makes money, information’s free, but they find a way to commodify it, and they make millions of dollars off of essentially the free information of the world. I kind of have a beef about libraries, say, uploading OPAC records to Google, because Google probably would make some money off of that or something. (P1a)

The perception of search engines as commercial and therefore at least somewhat untrustworthy is thought to be widespread by P9: “I think that tends to be a fairly common perception, certainly the people that I’ve talked to, within my library and in other libraries.”

4.3.1.2 Visibility as an Impossible Goal

Another attitude closely linked with that perception of libraries’ insignificance in the eyes of search engines is the sense that making library resources truly visible in SERP is not possible. There is a great deal of competition for those coveted spots on the first page of search engine results, and some participants think that library resources are unlikely to ever rise to the top:

> I think it would be naïve to expect a lot of traffic to be coming to bibliographic records from people searching in Google. Libraries just aren’t going to be the top result in a Google search if someone types in the title of a book, unless it’s an extremely obscure book that Google can’t find anywhere else. (P6)

> I think the problem for us is going to be that, as libraries, I don't really see how an individual catalogue record is going to end up at the top. (P3)

If true visibility requires rising to the first page of SERP, P6 is not optimistic about libraries' chances:

> There's just no way we're going to rank on that. You're going to get book reviews, you're going to get Amazon links, you're going to get the publisher's link, you're going to get all kinds of articles about it from different sources, but you're not going to get the X public library copy, even though our record is indexed somewhere in Google. It's just not going to come up high enough in a general search for that title. Unless someone actually searches "Go Set A Watchman, X public library", which I think is just not what people are inclined to do. (P6)

P6's library is one of the few Canadian public libraries to succeed in having the contents of its OPAC indexed by search engines in a meaningful way, and has done so for several years. This makes P6 the participant most qualified to discuss the impact, or lack thereof, of being indexed. P6's views are confirmed by website traffic data (see Section 4.6.2). In discussing linked data as a possible solution to libraries' visibility needs, P7 compares search engine visibility initiatives such as Libhub to early attempts at cataloguing the Internet in the 1990s:

> We wanted to catalogue the Internet. That was how we were going to be relevant, and raise our professional profile and stay in business, was because we were going to catalogue the internet. That didn't happen! That was overly ambitious. […] It really took a bunch of engineers at Google to do that. And so, I am excited about linked data, and I'm excited about making our stuff more discoverable, but I'm thinking, oh, you know, is it really… or is it going to be the next thing where we say, "Oops!" (P7)

Linked data and the Semantic Web, as the next major iteration of the web, present new unknowns, and new potential pitfalls. This pessimism may dissuade libraries from undertaking SEO initiatives.

4.3.1.3 Unknown Consequences

The flip-side to the feelings of pessimism described above is a feeling of trepidation about the results of a successful visibility initiative. Visibility initiatives present many unknown variables, both in their execution and their impact on existing library services. Some participants expect linked data initiatives like Libhub to have unknown and powerful effects on the library landscape. For instance, a truly successful visibility initiative might fundamentally change the dynamics between libraries and vendors:

> When you look at Libhub, which is the BIBFRAME project, one of the FAQs they have is, "Does this replace my discovery layer?" What they say is, "No, you still need it". But I think, well if everybody was really using that, then the search part of discovery layers would be a lot less important. What you would basically be looking for is just landing pages for individual things. In which case you might not even need a discovery layer. […] That's a lot of money not going into the ILS discovery layer industry. (P3)

Participants also expect visibility initiatives to have unknown impacts on the demand for library services:

> Exactly, we don't know yet. I mean I think the best case scenario is, "Oh wow, the library has it! Never thought of going to the library." I think that is what we want, but it might be, "Oh, they have it over there, I want to get my hands on it. How do I do that?" And then they might go to their local library and ask for it through interlibrary loan. (P8a)

Increased use of interlibrary loan is positive, and "has the potential to be very meaningful" (P8a), but also has the potential to be "a costly exercise" (P8a), and its eventual ramifications are unknown. When asked about any challenges experienced so far in the course of implementing her library's visibility initiative, P8a replied that she expects " some unforeseen" challenges to unfold as they move forward. The unknown effects of visibility initiatives may be daunting for some libraries, especially in terms of unknown costs and potential strain placed on existing library services.

4.3.1.4 Lack of Motivation to Become More Visible

Some participants express general support for the idea of making library resources more visible on search engines, but feel little urgency to approach the issue within their own libraries. According to P3, SEO is less urgent for her library because they don't particularly need the exposure:

> Our market penetration in terms of the number, the percentage of our community who are card holders is really quite high. So I think there's maybe less focus on trying to catch people with say an individual title. (P3)

P9 notes that in the absence of a real crisis to motivate his library, SEO is unlikely to be a priority:

> This seems like the kind of thing that would probably be driven by a crisis of some kind. So when I talked about some of the other libraries sort of feeling a sense of urgency about getting their stuff more visible to people out there, […] I think that's the kind of thing that could lead to a wider engagement and a need to pull together some of this stuff as more of a strategy on how to get our stuff more out there. (P9)

P1b identifies this sense of not needing more exposure as being specific to libraries. Having spent years as a web designer in the commercial world, he noticed a distinct difference in attitudes when he entered the LIS realm:

> In that commercial world, it definitely came up a lot more often. I find that here it's kind of an afterthought. […] I think people just assume here that people will just come to our website because they go to library websites. They don't think that we need to attract people that aren't necessarily thinking about the library when they're trying to get that book from Chapters or something, right? So yeah, it's definitely… I don't find it as much at the forefront of the discussion as it was in the corporate world. (P1b).

Libraries are a "community beacon" (P8b): prominent public institutions that may not feel a strong need to promote themselves or their services. P1b observes that this status can cause decision-makers to overlook the benefits of SEO:

> I think attitudes of managers, people in the library system, they just feel that we're a big huge public resource and people just will know to go there. They don't feel they have to sort of reach for the extra. (P1b)

As a result, simply having a library website that is accessible is seen as enough, without worrying about surfacing more detailed content:

> I think it's just seen as one of those Field of Dreams – if you build it, they'll come anyways. Which, to a degree, some do… but I still think we need to build up our presence there more. (P1b)

4.3.1.5 Concerns About Data Privacy and Security

Protection of privacy is one of the core values of librarianship (Canadian Library Association, 2015); concern over patron privacy and data security may present a barrier to librarians' willingness to undertake search engine visibility initiatives. In some cases, libraries are administered as branches of local government, which creates even stricter privacy and security restrictions:

> As a branch of government, we have some pretty strict regulations we have to abide by, so we do keep information that's not personally identifiable or even potentially identifiable. (P5)

At first glance, these concerns may not seem relevant to SEO. After all, the bibliographic records contained in libraries' OPACs do not generally contain any personal information, other than information about authors which is already publicly available. Discovery layers with social functionality may contain patrons' usernames in association with user reviews, but usernames are usually nonspecific and unlikely to lead to the personal identification of any individuals. However, this concern for privacy and data security manifests in a general wariness about opening up library data to outside sources: "So the barrier would be a desire for a secure environment and the limits that puts on people being creative about how they implement a solution for us" (P7). Librarians must be extremely cautious if there is the slightest chance of exposing information that patrons would prefer to keep private. P7 is cautious about allowing other organizations or

companies access to OPAC data without the presence of "a direct contractual relationship" that has been tailored to the library's specific needs:

> Many of these companies who provide these services aren't actually interested in locally customized contracts. They're interested in you adopting their standard contract. And you know, a standard contract just wouldn't give me the confidence in the IT security perspective. (P7)

Libraries work hard to address patrons' privacy concerns, resulting in caution that can also extend to well-known vendors in the library community, such as BiblioCommons:

> We do use BiblioCommons as our primary discovery layer, [but] we continue to offer our OPAC, which is a conscious choice because it's totally locally hosted. A very small number of people don't wish to participate in BiblioCommons because it's a third-party product, […] and they would prefer that they not have to give their library card pin to another third party, they want to maintain a relationship just with the X public library, not with a third party in addition to that. So we have the two. (P7)

When one puts together this caution surrounding privacy with the distrust of search engines as commercial entities discussed above, a serious barrier to visibility may arise. This is further shown by some participants' discomfort with the use of tools such as Google Analytics to collect aggregated data about website traffic usage. Some participants, such as P4, are quite comfortable with Google Analytics:

> There's no confidential information stored there, so there's no way to identify anybody. It's all 100% anonymous. If we could identify people we wouldn't be doing that, because we don't want that stored in a database somewhere. So it's just general traffic that you're looking at, it's very useful… (P4)

However, others have less confidence in the fact that patron privacy is not threatened:

> We're not really under any illusions that Google is not making use of that data in other ways, right, that benefit them of course; otherwise it wouldn't exist. […] So they're swallowing all of that stuff up and God knows what they're doing with it. (P9)

As a result, there is tension between collecting data and protecting privacy:

> We've really been wrestling with these two: the question of data, and the amount of stuff that we have, and then just the flip side of it, making sure that peoples' privacy is being respected. (P9)

In L1's case, use of Google Analytics was halted entirely for a short period due to privacy concerns:

> A couple of years ago, they found out that we were using Google Analytics and all of a sudden we were told to take it off the website because we weren't supposed to be providing any patron information to any organization outside of Canada. (P1b)

While Google Analytics was reinstated eventually, P1b's story shows serious concerns among librarians about search engines and library data. Library staff who are concerned about using tools such as Google Analytics may be even more unlikely to feel comfortable allowing OPAC data to be crawled by search engines.

Initiatives such as Libhub may help to allay these concerns by placing an added layer of distance between search engines and library data:

> The basic premise is privacy as a positive sum. So I think that's where just with the BIBFRAME project and that being sort of a separate layer, there's a value to that for us because it's sort of an added layer of privacy and security to make sure that there's no, the only data that's getting transferred over is bibliographic data, there is no access to borrower information, there's no customer information in that dataset at all. (P8b)

Concerns about privacy and data security are an important part of libraries' roles as curators and protectors of information for their communities; however, these concerns add another thread of complexity to the already complex conversation pertaining to search engine visibility.

4.3.1.6 Investment in Current Practices and Formats

Libraries have been using MARC to format their bibliographic records for a long time, and P7 identifies this long-term investment as something that may impact librarians' willingness to take on visibility initiatives:

> One of the things that we're very conscious of, is […] we have a huge amount of intellectual capital invested in the MARC record. Which doesn't mean that we shouldn't move forward with the times, but it does mean that we shouldn't abandon our MARC records. […] We have this huge investment in having created a very, very rich, descriptive bibliographic database, and what we're looking really forward to is being able to leverage that, and perhaps migrate it, translate it, develop it, whatever. But I think that's what holds a lot of people back, is they can't – it doesn't seem truly appropriate to just start over when we've got such a huge investment of intellectual capital there.

While increasing search engine visibility does not necessarily entail abandoning them, MARC records are perceived as a barrier to visibility (see Section 4.3.3.3). Even though attempting to improve a library's search engine visibility does not need to involve throwing "the baby out with the bathwater" (P7), the perception that moving beyond MARC is a requirement for becoming more visible may dissuade librarians from undertaking visibility initiatives.

P8b also identifies attachment to more traditional forms of outreach as a barrier to gaining buy-in for visibility initiatives: "for so many of our folks, they are absolutely convinced that what is proven is it is the poster on the bulletin board that is the beginning and end of the promotional work that we need to do."

4.3.2 Organizational Issues

Organizational barriers may make it difficult for libraries to prioritize search engine visibility, or even to be aware that search engine visibility is an area in need of attention. Without support from organizational leaders, visibility initiatives are unlikely to be

undertaken; without buy-in at all levels of the organization, any initiative undertaken may face serious challenges in implementation. A general lack of comfort with, and understanding of, search engine technology throughout organizations may make it more difficult to gain the support needed for visibility initiatives. Finally, lack of policy or guidelines surrounding search engine visibility may reduce libraries' ability to manage search engine visibility over time and assess the success of visibility initiatives.

4.3.2.1 Lack of Integration into Organizational Goals

Search engine visibility and SEO for library resources is not always seen as a high priority for libraries. In fact, in some cases, it is not on the list of priorities at all. Each library in this study has a different level of focus on this issue, ranging from very little awareness, to active pursuit of SEO initiatives within the organization. For a discussion of libraries that have actively integrated search engine visibility into their organizational goals, please see Section 4.4.3.

If SEO for library resources is not integrated into an organization's higher-level goals and strategies, such initiatives are unlikely to be allocated the resources necessary for success, if they are undertaken at all:

> In libraries it's still considered optional, and I think that it's probably not even considered often at the really high levels where you start setting priorities for the organization as a whole, which then determines where you put resources, and everything takes resources. (P3)

Some participants report that SEO in general is not seen as a high-priority issue in their organizations. P5 characterized her organization's interaction with search engines as "largely passive. We only specifically interact when there's a problem." Even if SEO for library resources is seen as a goal, within their own departments at least, it has often not yet been integrated into the organization's goals in any concrete way:

> So certainly it's a goal of our department. I'm not sure I'd say it's a strong goal of the library as a whole. Mostly because there's not a huge level of understanding of what the issue is. (P9)

> If you were to say that "for this organization it's in the planning stages", it's not – it's just in the planning stages in someone's mind of the organization. (P4)

This can be a daunting barrier to getting SEO initiatives off the ground. P1b described the difficulty in including SEO concerns in decision-making during the process of choosing a new OPAC vendor:

> Personally, I'm pushing for that, but I can tell you right now that I think it's going to be a hard sell with the higher-level ups or managers […]. Just because I think there's still an attitude of, you know, "we opened a huge branch […], that's what gets us more people to our website, more traffic, and so on." And it does do it, to a degree, but we still need that search piece. (P1b)

Without a place in the organization's goals, SEO can be submerged under many other competing priorities: "Unless it's a priority you don't spend a lot of time on it, because there's other things to do" (P1b).

This lack of priority, and therefore lack of time, can pose a barrier to creating a strategy for search engine visibility:

> So there's really a lack of time to devote to thinking about these bigger issues and really developing strategies around them. So I would say that's probably a bit of a barrier, and it helps to explain a little bit why libraries allow vendors to take the lead on doing some of this thinking for them. (P9)

P6 describes a lack of time and focus as adding to the difficulty of addressing other, more specific technical barriers: "It's not even that we haven't come up with a solution, it's that we haven't even had the opportunity. It would have to be sort of a project I think to look into what the right solution would be and how to address that" (P6).

4.3.2.2 Lack of Comfort With Technology Throughout Organizations

Several participants ascribe the absence of SEO from organizational goals, at least in part, to a general lack of comfort with technology among individuals in senior management roles within libraries:

> I think there's often a lack of understanding and real comfort with technology. So that when you talk about things like SEO or search engine optimization, that it's just not a concept that is clearly understood and the value isn't always understood, so when it comes down to setting priorities, if something is less understood, it either […] becomes a mythology where nobody says no because they don't want to sound stupid by saying no to something that sounds technical and important, or it gets paid lip-service but it never really makes it to the top because nobody understands the full implications. (P3)

While the participants themselves may be quite conversant with the relevant technology, they notice a lack of comfort among their colleagues. This lack of comfort with technology is described at all levels of organizations, and is identified as a barrier by multiple participants:

> I feel like there's sometimes misinformation, or lack of information, and a lot of lack of awareness in general among librarians. So it would be great to try to bridge that gap a bit and make people a bit more aware of just the basics even of how the internet works, in terms of, I'm not sure how many people know how search engines crawl the internet and how they work. (P6)

> Even just with people on a very base level getting familiar with our new website sometimes is a bigger challenge than one would think. (P8b)

> I think one of our challenges here is getting our senior management team to understand the issues. And also getting more of the front-line people to really understand what it's about. (P9)

Lack of comfort with technology is sometimes thought to be a generational issue: "Some of it is going to be generational, and some of it is comfort zone, and some of it is just where people want to put our energy" (P8b); "Throughout the profession […] people under 40 in sort of system-wide management or director type positions are not common, although they're becoming more common and I think we'll see a shift" (P3).

4.3.2.3 Lack of Buy-In Throughout Organizations

Another, related barrier to increased search engine visibility is lack of buy-in throughout organizations. P4 feels confident that buy-in will not be an issue: "I think we'll have the buy-in we'll need; I'm not concerned about that at all". Nevertheless, other participants express concern with library staff buy-in: "the only other barrier like I said is just people not seeing the need for it, I think, and trying to convince them otherwise" (P1b). Even L8, which has a high level of buy-in at the top levels of the organization, still struggles with gaining buy-in among front-line employees:

> I think in part the buy-in challenge may be that people don't necessarily… for people who are immersed in the organization, it's not necessarily a priority for them. For us, if we look at our staffing complement, a huge component of our staffing complement is dedicated to our frontline service, and the service that we offer in our physical buildings. So for them, their priority is that customer that's standing in front of them. And sometimes it can be a bit of a hard sell, or just a conversation that we need to invest some more time and energy into. (P8b)

It can be difficult to explain the connection between SEO and a library's front-line operations: "I think that the invisible nature of this work for many folks makes it a harder sell" (P8b). While front-line staff may not be making decisions about organization-level priorities, they are still perceived as having an impact on SEO initiatives:

> A lot of the people that work in our front lines, they're the ones that are having day to day conversations with library users, and if they're not aware of some of these issues, I suppose that can be a bit of a barrier as well to creating awareness out there. (P9)

4.3.2.4 Lack of Integration into Departmental Structures

SEO, which encompasses not only the library's main website, but also the library's resources, is further complicated by its multi-departmental nature. L2's participation in the Libhub Initiative, for instance, involves multiple different departments:

> We have a lot of people coming from Collections Management and Access because it's mainly about linked data and servicing your collection. And so we've got me in my new position, and we also have the cataloguing librarians because they've got to figure out how to transition, or how this new bibliographic framework could transition MARC, which is essentially our data in packages or silos, into linked data which would (and the initiative is, everyone is still trying to figure all this out), which would hopefully be better to work with all of the microdata from schema.org. So, it will all hopefully come together to help search engines find our data better and service our collections. And through that we have someone from Web Services sitting on it, because they're obviously going to have an interest in the microdata tagging and the linked data initiative aspects of it, and we have some people in our IT department sitting on it because they're going to have to think about how this is going to work with our integrated library system. (P2a)

Librarians are no strangers to inter-departmental cooperation, and this collaboration can in fact be fruitful: "At least in our organization, we talk to each other a lot and there are a lot of things we collaborate on that touch more than one area of service" (P5).

However, when there are no official and structured task-forces involved, the issue can fall between departmental cracks. In L1, for example, web development and SEO happen in a different department (and separate physical branch of the organization) than the development of the OPAC. Often, responsibility for search engine visibility of library resources does not fall into any staff member's job description:

> Where we don't have any staff dedicated to that, to doing web development as such, and application development, and search engines specifically, it kind of gets, that type of work kind of gets done when it needs to be done, or when

> there's free time. It's not a super high priority in some cases, so it could go months without being checked. (P10)

Often the participant to whom I was directed within each organization, as being the most qualified to speak about search engine visibility, did not actually have search engine visibility in their job description: "I think this is not, it's not even my official responsibility to be honest, I don't think it's on anyone's radar" (P6). Not having any staff members dedicated to this issue, or even any staff members who encompass this issue within their job descriptions, may limit an organization's ability to address the issue.

This lack of official integration into departmental structures also has an impact on communication between organizations. If SEO does not fall into anyone's official job description, it can be difficult to know whom to approach for more information about an organization's SEO experiences. This difficulty was encountered frequently in recruiting participants for this study. For example, "When you sent that email to the library, it literally went to 5 people before it got to me, and it just, this could have went straight to me" (P1a). Not knowing whom to approach can also hinder librarians' effort to reach out and share information between organizations:

> It's one of those things where it's hard to figure out, and obviously you've been going through this process, but it's sort of hard to figure out who within any given library is actually responsible for this, if anyone, and who you would talk to if you wanted to share information. (P6)

Since inter-library cooperation has been identified as a major potential enabler (see Section 4.4.1), this difficulty in knowing whom to approach may constitute a serious barrier for libraries.

4.3.2.5 Lack of Policy

As search engine visibility lacks coverage within organizations in terms of departmental roles and individual responsibility, it is hardly surprising that there is also a lack of policy

on the issue. None of the ten libraries participating in this study have any explicit policies for search engine visibility for library resources. In some cases, participants discuss tangentially-related policies, such as those related to not limiting patrons' choice of search engine within the library (P7), or general IT policies for internet use within the library (P1a). However, there are no policies that specifically address SEO in any of the libraries.

This lack of policies may present a barrier to visibility on search engines. Without consistent policies, even if SEO initiatives are undertaken, they will depend on a group of specific employees, and are likely to flounder if those employees leave the organization. In addition, without specific policies, SEO initiatives will likely lack incentives and metrics for success.

Some participants suggest that policy might be crafted eventually, but that it is too soon for policy at this point. As P4 says, "we're not there yet." P9 does not see the creation of policy as necessary at this point: "I can't really think of a way that policy would really help us, would really get us any further along than we are." Similarly, P10 has not found policy necessary yet: "We haven't had a need to do it, a policy for it." However, P8b can envision a future where policy might be necessary: "I can definitely see a future where there is going to be that." P3 sees not only a role, but a strong need for policy in the future:

> I think in 5-10 years it could be a very different picture. In the same way that 5-10 years ago we didn't have policies on social media use, and now that's just standard, right? So I would say, I hope that in 5-10 years we'll all have things on search engines and discoverability and indexing. If we don't, I'm a little afraid. (P3)

P10 identifies the rapidly changing nature of search engines as one of the barriers to creating policy: "It's a technology that changes so much that your best practices are always changing" (P10). P8b also sees these rapid changes as a policy challenge: "There are so many changes, search engines are reworking and changing themselves in so many different ways on an ongoing basis. How we keep up and where the policy line is in terms

of defining our involvement as it relates to that." However, she notes that libraries have experience evolving policies over time to encompass changing technology:

> So we sort of expanded the notion of the policy to try and pull in those various elements. I can see that applying to search engines as well over time as it crystalizes in terms of what the goal and intent are and particularly as you mentioned around the localization and the location features and how that relates to privacy, I think there probably will be a role for policy there. (P8b)

The general consensus seems to be that other barriers, such as lack of awareness and understanding of the issues throughout organizations, must be addressed before attempting to build policy for search engine visibility: "I'm not even sure that that base knowledge exists across the organization at the levels that it needs to. So I think I would even start at a more basic level" (P3).

4.3.3 Technical Difficulties

It is now technologically possible to open up library records to search engines. However, doing so is not necessarily easy. The themes identified here relate to two different types of technical difficulty: difficulty in having resources indexed by search engines, and difficulty in having resources reach a high enough ranking in SERP to actually be visible to a casual searcher. If a library resource is indexed, but only shows up on the twentieth page of results for a query, it is unlikely to be seen by many searchers. The following sections describe technological barriers as perceived by participants. While the perception of these barriers is real, and has an impact in and of itself, that perception is not always supported by fact. In this section, barriers are reported as perceived by participants; please see Section 5.2.3.2 for a discussion of outside evidence concerning the accuracy of these perceptions.

4.3.3.1 Difficulties Being Indexed

The first set of barriers discussed here centre on allowing library records to be indexed by search engines. 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). Stiller and Szymanski (2008) write on Google's official Webmaster Central Blog:

> While static URLs might have a slight advantage in terms of clickthrough rates because users can easily read the URLs, the decision to use database-driven websites does not imply a significant disadvantage in terms of indexing and ranking.

Based on this statement, search engines now have the technical capacity to crawl any OPAC, with or without stable URLs.

Nevertheless, having dynamic URLs can still pose a functional barrier to being indexed, if only because most librarians believe that dynamic URLs cannot be indexed:

> Everything is completely dynamic. The biggest hurdle with it is that it creates a, it carries over a unique session in the URL all the time. So every time a search engine tries to index the catalogue it's a new URL every time. (P10)

> It was a problem with the OPAC that X public library used before we created this new website, that the URLS contained session-strings and were unstable. And that is a massive problem. That's basically a non-starter if you're trying to get indexed in search engines. Stable URLS is just the table-stakes. You have to have stable URLs to be able to be indexed in search engines. If the crawler goes back and can't get the same result when it tries to revisit a URL, you're toast. (P6)

This perception, which is quite widespread not only among library professionals but among internet users in general, can cause libraries to block crawlers from indexing their

OPAC using /robots.txt files, even though search engines have evolved ways of dealing with dynamic content.

Libraries can also be forced to block search engine crawlers from OPACs because of the perception that the technology used in the catalogue is unable to withstand the process of being crawled. Some older systems may not be "mature enough to handle it, to handle being crawled" (P3), and being indexed by crawlers can have "performance-impact in the OPAC software" (P5). This again can lead to a /robots.txt file that blocks all crawler access, essentially ending any chance of search engine visibility.

Even if a library's OPAC does have stable URLs, in some cases librarians are hesitant to focus on getting it indexed until the catalogue has not only stable URLs, but *permanent* URLs. Having stable URLs means making sure that each item is attached to one URL which will persist from one search to the next. Having permanent URLs 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, no matter how often library records are updated and changed. For example, L2's records do have stable URLs; each item has one simple URL, rather than appearing on a dynamically-generated page created by each individual query. However, having *permanent* URLs is a further challenge. The crux of the problem, according to P2a, is how to deal with a very high volume of records, which are constantly changing. Books come and go from a catalogue:

> Say we own a copy of the book Maddaddam by Margaret Atwood. This book would have a webpage in our OPAC and ideally this webpage would be crawled and indexed by search engines. But, if we lose all our copies of this book we lose the corresponding OPAC webpage. Even if we get more copies at a later time, the new books will be assigned a new OPAC webpage with a different URL. So, if someone Googles Maddaddam and they are directed to the original OPAC webpage that user will receive a 404 error. (P2a)

In her opinion, libraries are working to "achieve permanent URLs that can be sustained through the additions, modifications, and deletions that always occur in a library's collection and catalogue," 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" (P2a).

The idea of having any dead links or 404 errors connected to the library is a worrying scenario for participants:

> If we're promoting that we have a resource in Libhub, but if that item no longer exists in our collection, when a customer goes to our catalogue [via Libhub], they could potentially get a 404 error. That's something we would like to try to avoid. (P2b)

> So every URL if it ever gets indexed is always unique so it only ever shows up once. So if we were to get indexed, someone clicked on it, the URL is going to bounce. (P10)

This perception can cause libraries to hesitate before opening up their OPACS, and focus on achieving permanent URLs first: "I think that permanent URLs are the ones that certain organizations are striving towards over site indexing" (P2a).

Another perceived barrier to having the contents of OPACs systematically indexed is a lack of links to individual bibliographic records. Links between pages are fundamental to the way search engine crawlers operate. Crawlers follow links in order to discover new pages, and when the trail of links dries up, the crawler turns its attention elsewhere (Google, n.d.a). Participants perceive this way of discovering pages to be incompatible with the structure of many OPACs, which contain bibliographic records as separate, unconnected pages called up in response to searches. As a result, participants believe that search engine crawlers gain access to these records in a patchy, inefficient way:

> Unless we submit some kind of sitemap or something, the search pages are the only avenue for, the only avenue to most of our records. The records are the ultimate thing we want indexed. And if we… most of our records aren't linked in any other way, other than through search, so we have to let the crawler crawl the search pages to get to the records. Until we come up with another solution. (P6)

> I don't know that those search engines have a way to meaningfully walk the database. I suspect that results could more commonly come from starting on the webpage with featured titles and things like that and finding the catalogue that way. So there's certain links that could get followed and then at a certain point they're going to be stopped, right, because there's nothing else to follow. (P7)

There are ways of mitigating this issue, such as sitemaps, but given the size and rapidly changing nature of most OPACs, these may be difficult to maintain.

Even if crawlers do gain complete, meaningful access to a library's OPAC, this may not lead to a complete indexing of all the resources the library has to offer. P2a, P2b, P3, P9, and P10 expressed concern over so-called "silos of data" (P2a). As P3 says, simply surfacing bibliographic resources is not enough; it is essential to surface the library's full range of resources and services:

> I could talk about discoverability of catalogue records, but I really think it has to be a bigger picture than that. I mean, this is about your entire web presence. It wouldn't make sense to have really great catalogue records come up but no programs come up, or no other information. Everything kind of needs to. So I think it would have to be a more holistic approach, and I do think that's more how libraries are going. […] Let's not just look at one silo of information or one silo of service in the library, let's look at the whole integrated picture of what we do. (P3)

This is a problem because so many of these other items, such as e-books provided through Overdrive, or magazines provided through Zinio, may not be integrated into libraries' OPACs:

> Even if BiblioCommons gets all of their stuff in there, and more visible, we're still missing a whole bunch of other stuff that we have. So, you know, I think in that sense, the fact that we have various silos of resources is a bit of a barrier to making this really work well. (P9)

Efforts to surface resources listed in OPACs are valuable, but some participants feel that true visibility for library resources must include "all of the different types of resources that we subscribe to through databases, or collections of electronic resources, ebooks, e-audiobooks, and so on and so forth" (P7). This may involve whole new sets of vendor relationships and barriers. Please see Section 4.3.5 for more discussion of vendors as a barrier to visibility.

4.3.3.2 Difficulties Rising to the Top of SERPs

Once library resources are being successfully indexed by search engines, the question then becomes how to ensure they rise high enough in results to be truly visible. L6 is one of the few libraries to have mostly complete indexing of library resources. While search engines are overall the most popular channel through which users find L6's website, that traffic is mostly driven through more general searches for library hours, library branches, and so forth (See Section 4.6.2). According to P6, L6 is quite visible in searches involving general library-related keywords, but does not see a high volume of users being brought to the library's website through searches for specific resources:

> The searches that I see that are driving traffic to our site are not… as I said, if you go deep down you will find searches for specific book titles, that have driven traffic to our website, but they're sort of way down in the results. Like, tiny. I dug around and saw that one of them was "Canadian Fundamentals of Nursing, 5th Edition". If it's a specific enough search, and there's not that many copies of this book or not that much information about this book available elsewhere… Someone did end up clicking through from a Google search for that specific book to our website, but that was like one person. (P6)

When it comes to more popular book titles, P6 feels that L6's resources are not rising high enough in SERP to be noticed:

> One of the really popular books right now is *Go Set A Watchman*, […] so obviously people are massively searching that in Google. But if you search that, even if you're in X city, you're not going to get X public library on the first or even the second or third page of your results. There's just no way

> we're going to rank on that. You're going to get book reviews, you're going to get Amazon links, you're going to get the publisher's link, you're going to get all kinds of articles about it from different sources, but you're not going to get the X public library copy, even though our record is indexed somewhere in Google. It's just not going to come up high enough in a general search for that title. (P6)

This is perceived as a problem because users are unlikely to go beyond the first few pages of SERP, and thus unlikely to encounter library content:

> I think it would be great to get library content on there, but I also think unless you're sort of on the first page, as a library I don't know that it's super worth it, because people only go sort of one, maybe two pages deep, unless they're really motivated. (P3)

Participants also report difficulty being visible in local results, even when items are being indexed by search engines:

> If you just had geolocation turned on and you searched, let's say 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. (P2a)

P6 and P9 also identify a lack of geolocation as a barrier to true visibility for library resources. Even when library resources do appear in the first few SERP, they are not always from the local library: "Right now in Google if you search for a title that pops up through a BiblioCommons link, it's not always necessarily the closest library to you" (P9).

Library resources can also have difficulty rising to the top of SERP if they are not presented according to SEO best practices as laid out by search engines. With large, constantly evolving databases, which were generally not constructed with search engines in mind, it can be difficult for libraries to conform to SEO best practices. For instance, the format of L6's catalogue is creating the flip side of the more common problem of not

having enough pages indexed; in L6's case, too many pages are being indexed. L6's catalogue provides faceted search, which allows users to refine searches by various criteria after the search query is placed. According to P6, this provides an important service to searchers within the catalogue by making search more intuitive and convenient, but creates a massive volume of URLs:

> Each of those refinements generates a different URL. The URL actually changes each time you refine. Which makes sense in a sense because every search is unique. But because we expose all of those refinements to people and to search engines, there are literally millions and millions of possible URLs on our website. (P6)

Search engines prefer websites with many similar or identical pages to identify canonical URLs (Stiller & Szymanski, 2008), in order to help crawlers understand which URLs are worth crawling, and which are unimportant duplicates. Although specifying canonical URLs would probably fix L6's problem (see Section 1.2.2.3), P6 says that her library has yet to begin addressing the issue: "We haven't come up with a solution to that. We just haven't. Or it's not even that we haven't come up with a solution, it's that we haven't even had the opportunity" (P6).

4.3.3.3 General Technical Difficulties

In some cases, the barriers perceived by participants are not connected specifically with either indexing or visibility, but rather relate to general preparedness for SEO initiatives. Sometimes these difficulties are not understood well enough to be precisely defined. For instance, in some cases participants have difficulty controlling their libraries' online presence in search engines, even down to the basics such as making sure the correct opening hours and phone numbers are visible to searchers. Some participants, such as P1a, P1b, and P4, describe tools like Google Plus business accounts or Google Places as being effective in controlling that type of information. However, others like P8b and P7 experience ongoing difficulty. P8b describes difficulties having the wrong information displayed in Google:

> We encountered another issue the other week where in this particular case Google was caching our hours and so they somehow cached some of our holiday hours, which then, when people doing Google searches, made it appear that some of our locations weren't open, when in fact they were. (P8b)

P7 describes difficulty controlling the order and appearance of basic branch information in the search results themselves:

> Our central library is announced on the website, our "Contact Us" is announced on the website, but for some reason when you Google us, the first thing that comes up is one of our branch libraries, and it happens to be a small branch library, and it happens that their particular individual page and their phone number come up in such a way that if you just looked at the search results, you'll see a phone number and you'll pick up the phone and dial the public library by looking at just the search result and recognizing there's a phone number there in the search result. And it's a real burden to us that people don't find our actual published main contact number but instead find the number of one of our small local branches who don't have the staffing to deal with the volume of calls that comes in. (P7)

Difficulty controlling the display of branch telephone numbers is persistent for L7, and causes a great deal of frustration: "We've encountered it more than once and it doesn't go away. It's like when they get the wrong answer, it's really hard to persuade them to stop" (P7). The lines of communication between Google and P7 are not open:

> Every mechanism that Google theoretically offers to allow us to have some input into the search ranking… doesn't work, doesn't work. They do not accept feedback in this area. Or if in fact they are honestly accepting feedback in this area, they are not able to fix it. Which might be, their algorithm just can't accept that correction, or it might be that they are vastly understaffed in the area of responding to peoples' complaints in this area, and I have no way of judging that because they don't respond to any comments, questions, demands for information. The only way to try and influence it is to fill in a web form. […] There's no way to talk to a person. (P7)

Google's machine-learning, web-form-based interface, scaled to serve millions of people all over the world, does not compute for P7; and P7's person-to-person, individualistic

communication style does not compute for Google. There seems to be a communication disconnect that is making it difficult for some libraries to control even the most fundamental aspects of their presence in search engines, let alone the visibility of resources:

> There had been pages that hadn't been archived properly, or things that had fallen through the cracks, that find their way to a search engine, and yet all of these things that we try and serve to a search engine on a silver platter are not necessarily getting the same attention that we would hope for. (P8b)

This can be discouraging: If simply controlling the appearance of branch hours and phone numbers is so difficult, how are libraries going to gain control over the visibility of hundreds of thousands of bibliographic records?

Another general technical difficulty perceived by participants is the current standard format for library records: MARC. For decades, libraries' bibliographic data has been contained in MARC records. This format is identified as a barrier to search engine visibility. Taking part in initiatives such as Libhub involves transforming MARC records into a new form such as BIBFRAME, which can be difficult:

> One of the challenges initially was evaluating the MARC to BIBFRAME mapping - ensuring that all of the most pertinent bibliographic data was being mapped properly and published accordingly on the Libhub site. (P2b)

The prospect of converting MARC records into something new is perceived as daunting:

> The bigger barrier is there is nothing now easily, straightforwardly available for us to magically map our MARC records into beautiful semantic markup that makes them wonderfully, immediately available through all search engines, totally discoverable. (P7)

> Adding another layer also means administration and management of that, and in a lot of cases, you know, also sort of managing the transfer of your files, your MARC, your metadata, to whatever form they need it in… (P3)

However, letting library records stay "locked into the MARC, these little boxes, these individual records" (P2a) is perceived as a barrier to visibility:

> We recognize that what we're challenged with in that regard is that the MARC records themselves, while they're open and available, aren't very discoverable because they don't really talk web language, yet. (P7)

4.3.4 Limited Resources

Libraries, like most organizations, operate with a finite amount of resources in terms of money, staff, and time. Limited resources are one of the most frequently identified barriers. Ten of the thirteen participants in this study (P1b, P2a, P3, P4, P6, P7, P8a, P8b, P9, and P10) cited limited resources as a major barrier to accomplishing or even attempting search engine visibility:

> The basic [barrier], which I'm sure you hear from everybody, is lack of resources, both money and staff, people to work on it. And sort of in terms of competing priorities. Obviously we all have people, so anything is technically possible, but when you have competing priorities you have to pick. (P3)

Because there is, as yet, no simple or straightforward solution to the problem of search engine visibility, it is difficult for libraries to assess how expensive or worthwhile it might be to take on such an initiative. The barrier of limited resources, when combined with the fact that the technology involved is not always well understood (see Section 4.3.2.2), leaves libraries unsure of exactly what needs to be done, but with the general impression that any option will be expensive. It is difficult to make decisions about resource allocation when the exact nature and amount of the resources is unknown:

> It's complicated, even what Zepheira and Libhub, what they're doing, it's hard to understand. So the conversation is just very high level, it's can we get our content there? And the answer is yes, for lots of work or money or resources or something. (P4)

In addition, there are no guarantees of success, and committing resources to a venture that might not pay off is a daunting prospect:

> Although we are interested in this area, we, you know, that whole time money thing. We wouldn't prioritize doing something ourselves. We wouldn't see that as a good use of our time over our many other initiatives, because it would be simply so all-consuming. And perhaps it would fail. I don't think it's a one-institution problem, frankly. (P7)

This element of uncertainty makes participants very hesitant to venture into the unknown by themselves. P7 is waiting for other organizations to develop a solution that her institution can adopt without committing as many resources, and without risking failure:

> We're waiting for others to be at a place where we can start adopting solutions that they have found or they have developed, and we're really looking forward to that. (P7)

This perspective is quite reasonable. However, if every library waits for every other library to break the trail, true search engine visibility for library resources may be a long time coming.

Even libraries that have made the initial investment in surfacing their collections, such as L6 and L8, still struggle with "competing projects and priorities and limited resources" (P6). In these cases, staff time is the resource that continues to be in short supply, even after an initial investment in infrastructure has been made:

> It would be interesting to have more time to spend on this, to spend on trying different things, experimenting and seeing if we could improve our visibility. So I guess in our case the main barrier has been just lack of time to focus on this stuff. (P6)

> Even though we want to be more visible on the web, we still need to keep all of our physical locations open. And some of the work that needs to be done is sort of increasing capacity when budgets are more in a holding-pattern and some of the day to day logistics haven't changed dramatically. It's not like we can sort of automatically add two people to our technology team that can

> really devote the time and energy that's needed to make some of this work as successful as it could be. (P8b)

Because SEO initiatives do involve significant investment and risk, having enough staff to assess their ongoing success is essential to determining whether there has been a sufficient return on investment. Without evidence of a real, positive impact, SEO initiatives may flounder and sink back to the bottom of the list of priorities.

4.3.5 Vendors

Vendors are an important player in the library world, especially when it comes to the technology used to manage library resources:

> Libraries – and I'm sure you've already heard this a lot – we've given away […] a lot of our control […] so that we can provide more powerful or more relevant or up-to-date resources. So this is one of those situations where we're kind of currently tied to what our vendors do. (P2a)

Of the ten libraries included in this study, eight rely on a vendor to provide their OPAC or discovery layer; all rely on vendors to provide their ILS. This means that vendors play a big role in the visibility, or lack thereof, of library resources. Participants frequently describe vendors as presenting a barrier to visibility, but many participants also look to vendors for possible solutions to the visibility problem (See Section 4.4.5). As an intermediary between libraries and the online display of bibliographic records, vendors can have a huge impact, both positive and negative, on visibility initiatives.

Vendors free libraries from the need to build and manage ILS, OPAC, and discovery layer software. Due to the limited resources discussed in Section 4.3.4, this has a strong appeal for libraries:

> There's really a lack of time to devote to thinking about these bigger issues and really developing strategies around them. So I would say that's probably a bit of a barrier, and it helps to explain a little bit why libraries allow vendors to take the lead on doing some of this thinking for them. (P9)

The price of this freedom, however, is a loss of direct control over the way bibliographic records are displayed online. Libraries have contractual agreements with vendors, allowing library input into how bibliographic records are displayed by vendor software; participants' perceptions, as reported here, do not necessarily reflect the exact nature of those agreements. However, based on these perceptions, it seems that the standard relationship between libraries and vendors does not usually allow librarians to directly modify the HTML used to display bibliographic records, and thus does not offer librarians the opportunity to experiment with the basic elements of SEO such as sitemaps, schema.org markup, and /robots.txt files.

Depending on what product is in use, in some cases the technology provided by the vendor is incompatible with search engine crawlers (see Section 4.3.3.1). In other cases, the technology does allow for search engine crawlers to index results, but librarians have very little control over how those results may appear to crawlers: "We don't have control over them submitting sitemaps, them dealing with /robots.txt files, them adding microdata to the catalogue" (P2a). Not having control over this technology can stop libraries from implementing any SEO strategy at all, especially if the technology provided is lacking in necessary functionality:

> There currently really isn't a strategy for the catalogue because our catalogue doesn't support a lot of the […] technologies that you'd use to improve your search engine rankings. We can't rewrite URLs, we can't add extra metadata, or keywords as such to the catalogue system to change it, you know? It's a proprietary system and very closed. And changes and documentation on how to make changes are not really documented very well, or forthcoming from the provider. (P10)

Any changes to how records are displayed, what types of metadata are used, and what type of SEO (if any) is performed require cooperation from the library's vendor: "One possible area of concern is that our more modern catalogue is a hosted third party solution, and it would require the cooperation of that host in order to implement things" (P5). This can pose a challenge to getting initiatives off the ground, because vendors may

not be interested in modifying their products to meet individual libraries' needs: "BiblioCommons in particular tends to operate on a consensus of the customers, rather than serving the needs of individual customers" (P5). If a library is using an older product, it becomes even harder to get changes implemented:

> They're not going to do much in terms of changing it. It's just more maintenance mode. They're not going to make any changes to help us improve that, unfortunately. (P10)

Lack of control, combined with a lack of responsiveness on the part of vendors, can pose a serious barrier to SEO: "It's something that would be more easy to address for an online catalogue that we had more control over individually, locally" (P5).

Even when there is a robust SEO initiative underway, vendors can be perceived as standing in the way of libraries' aspirations:

> In terms of making sure that our webpages are marked up with schema.org microdata, whether that's bib.schema.org or just the schema.org vocabulary, I hope that becomes ingrained, but because our catalogue is hosted by a third party, we can't really edit the html. So in terms of going forward, I hope the vendor community increasingly supports search engine optimization initiatives. (P2b)

Vendors often also control the type of data libraries can collect about the website traffic experienced by their OPACs or discovery layers. This in itself has a serious impact on libraries' ability to undertake and manage SEO initiatives, because having access to high quality data is identified as an important enabler of visibility (See Section 4.4.3). P1a, P1b, and P10 were unable to provide any website traffic statistics on their OPACs for this study, because they do not have the option of collecting Google Analytics data for those domains. When asked about collecting data on website traffic, P10 responded: "Right now we do it for our website because we can arrange Google Analytics. We can't implement that or haven't been able to successfully implement that in our catalogue as of yet" (P10). The fact that most libraries use multiple vendors adds a layer of complexity when attempting to assess the impact of visibility initiatives:

> Following the path of the user from Libhub to our online catalogue to other activities within the catalogue is a bit difficult because these interfaces are provided by third-parties. (P2b)

Without the ability to easily modify the webpages used to display bibliographic records in order to apply SEO best practices, and without the ability to effectively monitor website traffic to OPAC records in order to assess the effectiveness of current SEO practices, libraries are left with few options when it comes to improving the visibility of their records.

The larger impact of vendors' role as intermediary is that libraries are slow to take independent action on technological issues such as SEO:

> That often happens in libraries, libraries tend to wait around and see what the vendors are going to do, rather than really press them to do things that are goals of the organization. One of the fatal flaws of all the libraries I've come in contact with. (P9)

P6 thinks this is a worrying trend:

> Yet again, instead of actually learning how the web works, and understanding how search engines work, it's like, "buy another product from a vendor." It worries me. (P6)

According to P6, having up-to-date SEO is not as complicated a problem as it may seem, but with vendors standing between libraries and their resources, libraries don't have the opportunity to discover that fact:

> It's not actually that hard to do it right. When you have access. The problem is that most libraries just don't even have access to touch their webpages, and they don't have access to the tools to understand how crawling works. That's the problem. It's not actually really complicated or hard. (P6)

The use of vendor-provided OPACs and discovery layers also poses a more concrete, technological barrier to search engine visibility. In most cases (including all eight libraries in this study that use vendor-supplied OPACs), when vendors supply a library with an OPAC or discovery layer, it comes with a separate domain name. This causes the library's online presence, and online reputation, to be split between two domains: the library's main website, and the OPAC domain. Participants are aware of this issue as a usability concern for patrons:

> When you're at, for instance, […] [another public library] website, […] and you would search for something, you'll actually leave the website. You will go into whatever product that they use – I think it's maybe BiblioCommons. So then you're not there anymore, right? There's that disconnect. You've sort of lost [another public library]. They've spent a lot of time trying to make it feel like you're still there, but you really aren't. (P4)

However, having a separate domain for OPAC records can also be an issue for SEO. L6 has invested the resources to build a unified, in-house domain that encompasses the library's entire online presence:

> All of our content is under our own top-level domain, like it's all www.xlibrary.ca, all these other libraries, they have, the actual domain is some type of vendor name, or something weird like Webcat, or BiblioCommons, or whatever … I don't know how much that impacts SEO, but it's not, it's obviously not ideal. (P6)

The specific factors considered by search engines' ranking algorithms are a closely guarded secret, in part to protect search engines from the malicious manipulation of their algorithms (Granka, 2010). However, it is known that these factors come together to allow search engines to judge not only the relevance, but also the reputation (either positive or negative) of each domain. When libraries use two separate domains, they are essentially splitting any positive reputation they may have acquired in two, leaving both domains with weaker reputations (Google, 2010).

The barriers posed by vendors discussed could all be described as unintended consequences of the library-vendor relationship. However, there is some doubt about

whether vendors are truly motivated to mitigate these issues, or whether the idea of library records being fully accessible through search engines would threaten vendors' role in the library community:

> There is a part of me that wonders if they worry that, if you could freely access everything in the catalogue through Google, then do you need a discovery layer? Or are you willing to pay as much for it? Or does it need to look different? Because the whole point of a discovery layer is to create a friendly, Google-y/Amazon-y interaction. So I sort of wonder in some ways if I were them, would I lock it down just because it's like, "Well, we don't want them getting the idea that they don't need us". (P3)

P9 also questions some vendors' motivation when it comes to making data more easily accessible:

> It really depends on a willingness between the vendors to provide that information, and in the case of Overdrive as an example, they really want people to come to their website. Because if all they are is a repository, the business model starts to fall apart a bit for them. (P9)

P3 and P9 perceive a potential conflict between the business model employed by vendors and the idea of making library resources easily accessible in online searches. This is not to suggest that vendors will not adapt, but if this conflict does exist, it may impact their level of commitment to SEO initiatives, at least in the short term.

4.4 Enablers of Visibility

While most participants were quick to identify current and potential barriers to search engine visibility, relatively few participants had perceptions to share regarding factors that might enable visibility. This hesitance may stem from the fact that the majority of participating libraries have yet to make any concerted effort to become more visible, or even to consider doing so. Participants from L8 and L6 had the most to say about enablers, perhaps because these libraries have been more active than most in pursuing the visibility as a goal, and thus have more first-hand experience of possible enablers.

4.4.1 Inter-library Cooperation

Communication and cooperation among libraries is identified as a much-needed enabler of search engine visibility for public libraries. The question of how to increase the visibility of library resources is seen as too complex for any one institution to solve alone. As P7 puts it, "I don't think it's a one-institution problem, frankly." Cooperation and information sharing are seen as one of libraries' strengths by participants: "That's what we do, we share!" (P8a). The area of search engine visibility is fertile ground for libraries' natural tendency to share:

> The beauty of being a public library, is that we're not competing with those other libraries in any real sense. So we can take advantage of what other libraries learn about things like this. (P9)

P10 also sees the potential of inter-library cooperation to solve libraries' visibility problems, but thinks libraries have some work to do in improving communication channels:

> The search engine thing is very important in the library, and it seems like there's a lot of people that are going in different directions with it. […] There'd probably be some value in having some type of online library community where everything is centralized, so people can go and share their expertise on search engines, for example, and how it relates to a specific product. (P10)

P6 would also like to see improved communication between libraries on this topic:

> I wish I had more people, I had more opportunities to talk about this with other people in libraries. As I mentioned, I've had a couple people from other libraries contact me in the last six months about SEO and search engine indexing and stuff. So there's obviously some growing interest in this area. And I think there's a need for more opportunities to share information about this. (P6)

Libraries have great potential to collaborate, but need to work harder to open channels of communication when it comes to search engine visibility: "Considering that libraries are

in the business of providing and sharing information, you'd think that it's an area that they would excel in the sharing of information between libraries, you know?" (P10).

Inter-library cooperation takes on a different form in the Libhub Initiative. According to P8b, part of what will make Libhub successful is the fact that it opens up a way for libraries to collaborate on SEO:

> So much of what libraries do, we work together. And there have been so many areas that we've been able to work together on, and search engine optimization up to this point really hasn't been one of them. So to be able to start to look at doing that bodes well for the future. (P8b)

Libhub harnesses that potential not only for information sharing, but to actually improve the ranking of library resources in SERP. The idea is that by banding together, libraries can combine their positive reputations by interlinking each bibliographic record they publish with records from other libraries with the same resource:

> So with the early adopter program, we publish our records, and we publish them to our URL that will be distinct to our library, but it's a subset of a URL for all of the libraries that are involved with the project. (P8b)

This is expected to "expedite the visibility of the record" (P8b) by pushing it further up in SERP rankings. P8b thinks this approach will be especially effective for public libraries, because they share so many records in common:

> In contrast to an academic institution, if you were to look across 50 public libraries, there's going to be a much greater percentage of the collection that we are going to have that is going to be similar or the same. (P8b)

Until now, public libraries have not necessarily been able to capitalize on this similarity of collections, but P8b thinks that Libhub will unlock that potential for collaboration:

> Because we share so much in terms of common resources that we make available to our community, it's not – what really excites me is that not only are we going to be visible in a different way on the web, but we're going to be able to work together in a way that public libraries traditionally haven't

> been able to as much, because we serve distinct communities. But in this case we can still continue to serve our distinct communities, but benefit from the fact that there are 20 public libraries across North America that have this in their collection, which means we have that many more records, which means we have that many more links, which means we have that many more hits. And we can really leverage that partnership to create a better user experience for the people within our community. (P8b)

4.4.2 Attitudes

If negative attitudes about search engines may pose a barrier to search engine visibility, it follows that positive attitudes may act as an enabler of visibility. Librarians who view search engines in a positive light may be more enthusiastic about undertaking visibility initiatives, and more open to learning about how search engines work. Despite the negative perceptions reported in Section 4.3.1, many participants also have positive things to say about search engines. Search engines are seen as powerful and convenient tools, popular with both patrons and library staff:

> They're just convenient. They're a part of our everyday lives. (P4)

> I think they're a wonderful resource. (P5)

> I use Google all the time in my personal and professional life. It's incredibly useful. (P2b)

> I know that they are very powerful tools that are incredibly relevant to our customers. (P2a)

> I'm an information professional, and that's where I get most of my non-fiction information, you know, unless I'm doing really professional research where I'm looking for scholarly journals or maybe something medical, but even then, these days, there's such good content, and the ease of use is so amazing. (P3)

This awareness of search engines' utility may encourage librarians to consider the effort involved in pursuing search engine visibility worthwhile. Seeing search engine visibility

as a worthwhile goal is necessary to the success or even existence of search engine visibility initiatives. The general usefulness of search engines, and their potential usefulness as a way of promoting library services, appear to be linked in P5's mind:

> I find them incredibly useful. And from the perspective of OPACs, and in general, I think the more exposure we can get, if that's through a search engine or any other exposure of the collection is to the benefit of the library and our potential users. (P5)

As noted in Section 4.3.1, participants do have reservations, and some, like P1a, would prefer not to open their data up to search engines at all. However, for the most part, participants are enthusiastic about the idea of search engine optimization, even if their libraries have not yet made any moves in that direction.

Participants highlight the possible benefits for libraries and their communities:

> If we were leveraging the search engines as much as possible, people would be able to […] just do a search for whatever title of a book they're looking for, and a link should come up that takes them directly to our catalogue, where they can place a hold or check a book out. (P10)

> I would just love us, for the catalogue, or really for the library, to actually be doing search engine optimization. (P3)

> I mean it's just the right thing to do to ensure that your data is discoverable to machines. (P6)

These positive views of search engine visibility are an important enabler of visibility; if visibility is not a goal librarians are excited about pursuing, it is unlikely to be attempted.

4.4.3 Organizational Enablers

P8b and P8a both cite a culture of innovation and flexibility within L8 as a major enabler of their library's LibHub early adopter initiative:

> And just creating that culture of innovation, which I think, as has been alluded to, it can be challenging because lots of people have lots of competing priorities and taxes on their daily schedule. But X Public Library has had a history of doing that stretch goal of wherever possible being an early adopter and so that sort of served this project well as well. (P8b)

> So I think that's sort of the reality of today, though, is […] the flexibility, the nimbleness, we're trying – we're in a situation now, all public libraries, where we're trying to get out there to reach the customer in different ways, whether it's outreach programs, or improving the website, or improving accessibility, it's all part of the same picture. (P8a)

P9 also sees organizational culture as a potential enabler, and cites Google as an example of creating time and space for experimentation:

> Certainly within my department, we try very hard to give our developers free time to experiment. Do sort of the Google thing, and allocate some time to pet projects, and things that maybe will not result in anything interesting, but just time to experiment. So I think ensuring that people have that time to just kind of work on what's interesting to them is important. Hard to find, but making it be a part of the culture, that's a bit of a challenge. (P9)

Innovation and experimentation often involve taking risks, which requires support at the highest levels of library management: "And recognizing the risk, that sometimes that means that maybe it's not going to turn out the way that it's anticipated, but that particularly in this instance it's worth the risk" (P8b). P8b identifies L8's library board as an essential player in supporting and investing in innovation:

> Definitely the library board. We've been very fortunate to have a board that has been very supportive of being future focused in terms of the library. I know that's a real gift because for this particular piece, for this project, we had to go to the board to ask to release funds from our reserve, and to try and explain web visibility to that group of individuals could very easily have gone a different way. But I think that support of recognizing the importance of technology and looking outside of our doors to make the inside of our doors more relevant, the board is absolutely key, particularly in a public library setting. (P8b)

4.4.4 Technological Enablers

Being well-situated technologically to take on a visibility initiative is seen by P8b as an important factor in enabling libraries to be more visible. P8b notes that a recent move to new ILS software had inadvertently laid some important groundwork for L8's Libhub early adopter visibility initiative:

> I think in a lot of ways we were very well positioned for the program. We migrated to a new ILS in the fall of 2013, so we had done, not that long ago in terms of library time, done quite a bit of extensive work on our ILS in terms of our authority control and database cleanup. (P8b)

The preparation necessary to migrate to a new ILS left L8 with clean, recently-vetted bibliographic data, which was easier to prepare for conversion into Libhub's BIBFRAME format:

> We had quite a streamlined process because we'd already… a lot of those question-mark fields or historical data that we might have forgotten about, in a lot of ways has already been dealt with, so that's facilitated our testing process for sure. (P8b)

Having up-to-date ILS technology is also an important enabler, as it frees up an all-important resource, staff time. P4 sees his library's ongoing ILS migration as a necessary first step before undertaking any search engine visibility initiative: "Everything we're doing now is to free up those resources, so that we're not just maintaining a system status-quo, so that we've got time and expertise to get to the next step" (P4). Increasing the library's ability to take on the problem of search engine visibility is certainly not the only factor in the decision to upgrade to a more modern ILS, but increasing L4's ability to innovate is definitely a goal:

> The [current] system's result is that we just spend so much time maintaining it and caring for it, and so we're moving to something that's going to free up some resources and really set the stage and give us a foundation for doing things that are more innovative. (P4)

Another important enabler is control over library technology. Direct control over the technology surrounding bibliographic records allows L6 to experiment freely:

> X public library is fortunate, that we have been able to have more direct control over this stuff than some other libraries have, so… I think that type of, that's been really important that we have direct control. (P6)

Having access to accurate data is also identified by several participants as an important enabler of SEO. Detailed website traffic data can help libraries focus their efforts in the most effective directions, as well as assess the impact of changes as they are made and adjust accordingly:

> And we also have visibility, that we are able to track things through Google Analytics and Webmaster Tools. I can't emphasize enough how important it is to at least, when you're embarking on a project, to try to understand your status in search engines, to have these tools to be able to see what's going on. I've talked to staff at other libraries who just don't have those types of tools available to them for their OPACs because they're managed through a vendor. I think that's something that vendors really need to enable for libraries. (P6)

P10, who currently does not have access to Google Analytics data for L10's OPAC domain, highlights gaining access to that data as an important step:

> Having Google Analytics implemented would give us way more information on our users, what they're searching for, and what's trending. Which is a real key factor because then you can take that information and adjust any policy or any goals that you have. (P10)

Accurate data is also essential to assessing the success of an initiative. L2 has been part of the Libhub early adopter program for over a year, and P2b hopes that taking a close look at the library's usage data will yield some important insights:

> The next step is to really analyze the data using Google Analytics and assessing pre- and post- involvement in terms of referral traffic, and also

> seeing […] if we can follow a user's path, so once they come from Libhub, land on our catalogue, see what activities we can analyze from there. (P2b)

4.4.5 Vendors

4.4.5.1 OPAC and Discovery Layer Vendors

Although vendors are frequently identified as a barrier to search engine visibility, they are also envisioned as potential enablers. As noted in Section 4.3.5, vendors are frequently seen as the party with the most active control over the display of bibliographic records online, and thus the party that must take action to make records more visible. As a result, librarians are looking to vendors for solutions: "Going forward, I hope the vendor community increasingly supports search engine optimization initiatives." (P2b). Although most vendors of library systems are not currently addressing search engine visibility for library resources effectively, that may change as more libraries emphasize the issue as a priority. P9 believes that the more libraries express interest in this issue, the more motivation vendors will have to address it:

> These vendors, they all know each other, they know what each other are up to, and when they start hearing their customers asking things, more than just a one-off… when they start getting lots of questions, I think they know they need to respond in some way.(P9)

This becomes an important avenue for the inter-library cooperation discussed in Section 4.4.1. In order to push vendors to enable search engine visibility, libraries must communicate and band together: "We'd have to convince people like New York Public that it's in their interests to push this as a priority as well" (P5).

BiblioCommons is a popular choice for discovery layer software among Canadian public libraries, and does offer the best overall visibility among the traditional library vendors used by participants of this study (see Section 4.5, Table 4). According to P9, BiblioCommons is responding to libraries' growing interest in search engine visibility by attacking the problem themselves:

> BiblioCommons is actively working on […] making […] the records available through their system more readable via search engines using the schema.org profiles for various creative works. (P9)

BiblioCommons has made some progress in promoting visibility for library resources, and P9 is hopeful that as search engine visibility becomes a more widely discussed topic, BiblioCommons will step up: "BiblioCommons has always been, we've always found them to be quite forward-thinking in a lot of these areas" (P9). However, it may take some time to see real progress:

> I know that in discussing with some of my colleagues at other libraries as well, the sense was that BiblioCommons tends to bite off more than they can chew. They tend to make fairly big promises sometimes, and under-deliver on those promises, or at least, I shouldn't say under-deliver, but sometimes it takes them more time. (P9)

Regardless of the shaky timeline, P9 is putting faith in his vendor: "BiblioCommons is certainly pretty key to our success in this area, because we have no real intentions at this point anyway of moving away from them" (P9).

4.4.5.2 Zepheira and the Libhub Initiative

The Libhub Initiative, run by the web technology company Zepheira, has positioned itself as an enabler of search engine visibility for library resources. Libhub is a test-case for the Library of Congress' BIBFRAME linked data model, which is an attempt to unlock the potential of linked data for libraries (See Section 2.2.4 for background on the initiative). As a vendor-run solution that libraries can purchase, Libhub appears to be the answer librarians have been anticipating from vendors. Two of the libraries in this study are participants in Libhub's early adopter program, and have positive things to say about Libhub so far: "'it's been a really positive experience for us" (P8b). P2b is also enthusiastic about her library's participation: "In terms of the visibility, yes, I think the initiative has been successful in making our collection more visible online," but is cautious about making definitive statements until the initiative's impact can be assessed:

"In terms of impact on whether that has increased circulation, or the number of library cardholders, I can't say at this time" (P2b).

Participants describe Libhub as transforming their libraries' collections:

> Libraries have such beautiful data. The cataloguers have really done a wonderful job of making things that are authoritative, and cohesive and comprehensive. But it's locked into MARC, these little boxes, these individual records. BIBFRAME will break those into individual data points, and then anything can be done with those individual data points. (P2a)

> And just the links, I think we recognize how that […] really transforms the collection in terms of points of entry and linkages. It's phenomenal how many points of contact 500,000 bibliographic records can metamorphose into. (P8b)

Libhub appears to be very centred on libraries' resources:

> For the first piece of the early-adopter initiative, the emphasis has really been on our traditional sort of collections and resources, not so much the database products or the other pieces. So I think it will, the first layer of impact will definitely be in relation to those borrowing collections. (P8b)

However, P2a believes that over time, Libhub will also help libraries address their online presence as a whole:

> They're also going to be looking at offering libraries help in servicing just their website information, getting you on the Google card, making sure that you're on […] Wikipedia, those sorts of things, so making your website readable by search engines. (P2a)

While the Libhub adopters in this study are excited about the initiative, other participants are not as enthusiastic. P7 is watching Libhub with interest, and is willing to be convinced, if not willing to jump on board immediately:

> So, really looking forward to what happens with Zepheira and other initiatives in that area, to start seeing the catalogue records and the

> components of the catalogue records be much more discoverable, to build linked data sets and collated information about something that includes the fact that the library has materials that can support peoples' discovery about say a particular author or particular topic or whatever. (P7)

However, others look at Libhub with skepticism. P6 is concerned that Libhub, as another added layer, distracts from vendors' responsibility to bring their underlying catalogue products up to the standards set by today's online environment:

> I mean, maybe I'm missing something, but that's... I guess what really sort of tipped me over the edge was seeing an announcement that Libhub is now partnering with SyrsiDynex. Because my immediate reaction was, okay, so SyrsiDynex is now going to sell libraries a Libhub add-on? Why don't they just fix their catalogue products? That's the problem! It's SyrsiDinex catalogue products, and other library vendor catalogue products that are the problem! (P6)

In P6's opinion, linked data and BIBFRAME, while important and valuable, may not be the answer to libraries' visibility problems:

> I think, I'm a little concerned that libraries are being led astray a little bit on this topic. I mean, the whole issue of […] how we catalogue and using MARC versus BIBFRAME versus something else, is a whole separate issue. And linked data is a really important concept, and it's really important for libraries to discuss moving beyond MARC, absolutely. But I guess when the conversation is framed around SEO and when libraries are being sold this idea of "get your records into Google!" that, you know, yet again, instead of actually learning how the web works, and understanding how search engines work, it's like, "buy another product from a vendor", it worries me. (P6)

P9 is also not convinced that Libhub is worth the price: "We declined to join their pilot, […] in part because it was a lot of money they were asking for, in relative terms" (P9). Based on conversations with L9's current OPAC vendor, BiblioCommons, "it seemed as though the whole Libhub Initiative was somewhat redundant to some of the work that they were already doing, especially as a BiblioCommons library" (P9). BiblioCommons is already fairly successful in getting library resources indexed by search engines, even if

they may not be rising as high as libraries might wish in SERP. This leaves P9 less than motivated to become a Libhub adopter:

> So it seemed to us redundant to jump on board with Libhub because […] if you look at results in search pages, using Google right now […] if you search for Gone Girl and [X Libhub adopter] public library or something, you get both of those results together in the same search result page. You get their BiblioCommons record, and you get their Libhub record. And when you go to the Libhub record it just links over. It brings you to a page, it stays there for about 10 seconds, and it refers over to a BiblioCommons record. (P9)

The added value of Libhub is not clear to P9: "it seemed to us kind of pointless to go through the effort and the cost of converting all of our MARC records into this other format, only to have them appear together with BiblioCommons and then refer back anyway" (P9). Libhub's current focus on bibliographic records is also a stumbling-block for P9:

> BIBFRAME and Libhub and that model, it's very focused on the Bib record, the work. […] But what strikes me as BiblioCommons' forward thinking on this, is that they're really focused as well on the whole social side of search too, more of the whole sharing culture. And in the webinar that I attended recently, they had, I don't know what their source was, I haven't really delved into it too much, but they talked about 30% I think they said of all referral traffic is through social media now. So you can't really ignore that, nor would you want to. But I think the BIBFRAME model is not really addressing that side of things, at least not as far as I'm aware. (P9)

P6 is also concerned by Libhub's biblio-centric model: "'It's definitely about more than books, more than about just MARC records and stuff like that. It's about the program info, it's about the branch info, it's about services, it's about everything" (P6). As noted in the quote from P2a above, the long-term plan for Libhub does include a broader focus on libraries' online presence as a whole. However, the fact that Libhub centres on the BIBFRAME data model does give it an undeniable, unavoidable focus on bibliographic records, which is unlikely to change in the near future.

4.5 Site-Specific Searches

The above results from the semi-structured interviews reveal the current landscape of search engine visibility in Canadian public libraries, as perceived by members of the Canadian LIS community. In order to more accurately assess those results, quantitative data was gathered about the participating libraries' current level of search engine visibility. Site-specific searches were performed in Google to assess the amount of pages indexed for each library's OPAC domain, as well as the quality of results retrieved. Searches were performed at roughly the same time as the interviews, in order to enable meaningful comparisons. In the case of L1 and L2, where two interviews were performed, the more recent site-specific search data is used in order to increase its comparability with the other libraries' data (see Section 3.3.2 for more details). Table 4 shows a summary of the number of pages indexed by Google for each library at the time of the site-specific search.

Table 4: Overview of library pages indexed by Google

Library	Number of Indexed OPAC Pages	Number of Indexed Libhub Pages	OPAC or Discovery Layer Vendor
1	481	Not a participant	Aquabrowser Library
2	214,000	72,200	BiblioCommons
3	269	Not a participant	Sirsidynix Enterprise
4	29,100	Not a participant	Own system
5	51,700	Not a participant	BiblioCommons
6*	7,370,000	Not a participant	Own system
7	221,000	Not a participant	BiblioCommons
8	207,000	204,000	BiblioCommons
9	213,000	Not a participant	BiblioCommons
10	1,410	Not a participant	Horizon Information Portal (Sirsidynix)

*Note: Because L6 does not have a separate OPAC domain, it was not possible to perform a site-specific search of its OPAC. This result is obtained through a site-specific search of L6's main (and only) website.

Based on amount and quality of search engine results, libraries have been divided into groups, ranging from libraries that block search engine crawlers entirely, to libraries that have many rich and informative results in their SERP. Please see Appendix G for screenshots showing a sample of SERP for each library's OPAC, and (where applicable) Libhub domains. In order to better understand the quality of results, each library's /robots.txt file was also examined (see Appendix H).

4.5.1 Libraries that Disallow Crawlers

In the case of L1, L3, and L10, the library's OPAC has a /robots.txt file that denies access to search engine crawlers. This means that all associated SERP look the same. They include the catalogue's title, and the following statement: "A description for this result is not available because of this site's /robots.txt". This causes the SERP to be non-existent in practice, making "zero" a more accurate estimate of indexed pages for each library. In the case of L3 and L10, all robots are disallowed from entering any part of the domain. In the case of L1, an examination of the /robots.txt file shows that all crawlers are disallowed in every part of the domain except "/sitemap.ashx". Directing crawlers to a sitemap is actually good SEO practice; however in this case the sitemap does not appear to exist. Following the specified URL path leads to a page that reads "Sitemap not found".

4.5.2 Libraries that Allow Crawlers Some Access

L5 appears to allow crawlers to index some, but not all, of its OPAC pages. Although at first results seem to be populated with detailed information, around the seventh page of results the titles revert to an unintelligible URL, and the snippets revert to the typical "A description for this result is not available because of this site's /robots.txt," which denotes a /robots.txt file which blocks crawlers. This means that Library 5's real number of

indexed results is much lower than the total shown. While SERP include resources from Library 5's catalogue, many results seem to be user profiles, user-created lists, and user comments. As individual user profiles are unlikely to be relevant in an average search, the value of these results is questionable. The title of each result includes the library's name, as well as more specific information about the resource. Each result also includes a snippet of information about the page, but often this snippet is not particularly intelligible. For example: "[username]. X Public Library Service. Message Follow Ignore · 0 Following | 0 Followers. Profile · Shelves · Lists," or "Subject Headings. No Subject Headings have been added to this title yet. Staff Lists that include this Title. No lists include this title yet".

L4 is somewhat difficult to assess because as a consortium, it has many different domains within its purview. I examined the domain connected with the catalogue shared between the consortium libraries, as the most reasonable equivalent to the other libraries' OPAC domains. However, because L4 is so different, it is difficult to make true comparisons. L4's search results do include some specific resources, but also seem to include many general search and navigation pages for the different libraries in the consortium. It is difficult to identify what the results are, because the majority of titles are simply "XILS", the acronym that corresponds with the consortium's name. This could be quite confusing to a searcher. The snippets included with results are sometimes intelligible, and sometimes difficult to understand. For instance, in Figure 3, the snippet has been taken from the table of contents of the book in question, giving few clues as to where the link will lead:

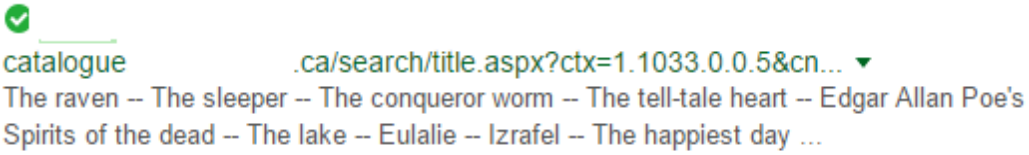


Figure 3. A rich snippet taken from L4's SERP

When combined with uninformative titles, these snippets may make L4's results difficult to navigate.

4.5.3 Libraries with the Majority of Content Indexed

In the case of L2, L7, L8, and L9, crawlers appear to have been allowed to index the majority of OPAC content. Results generally include informative titles and snippets, but the quality of these elements varies slightly between libraries. For example, L2's results include a specific title, followed by the title of the OPAC. The title does not include the name of the library, however, which could cause confusion to a casual searcher. L7's results, on the other hand, include a title with the library's name, the type of result (e.g., "Event"), and a more specific heading (e.g., "Holiday Puppet Show"), as well as a snippet that is usually quite intelligible (e.g., "Celebrate the season's winter holidays with a festive puppet story at the library. For ages 6 to 12. Program: Puppet Shows; Suitable for: School ..."). For these libraries, the results are a mix of resources, library events and services, user profiles, and user-created content.

While most of the results for these four libraries were quite intelligible, L2 and L9 had some results that might appear strange to a casual searcher. L2's results included a number of pages titled "MARC View", which contain MARC record data (see Figure 4). The MARC record found here contains the same information shown in the public view of records, but lacks the formatting that helps make bibliographic records intelligible to searchers. Seeing this page in a list of results for a search engine query might not entice a searcher to further explore library resources.

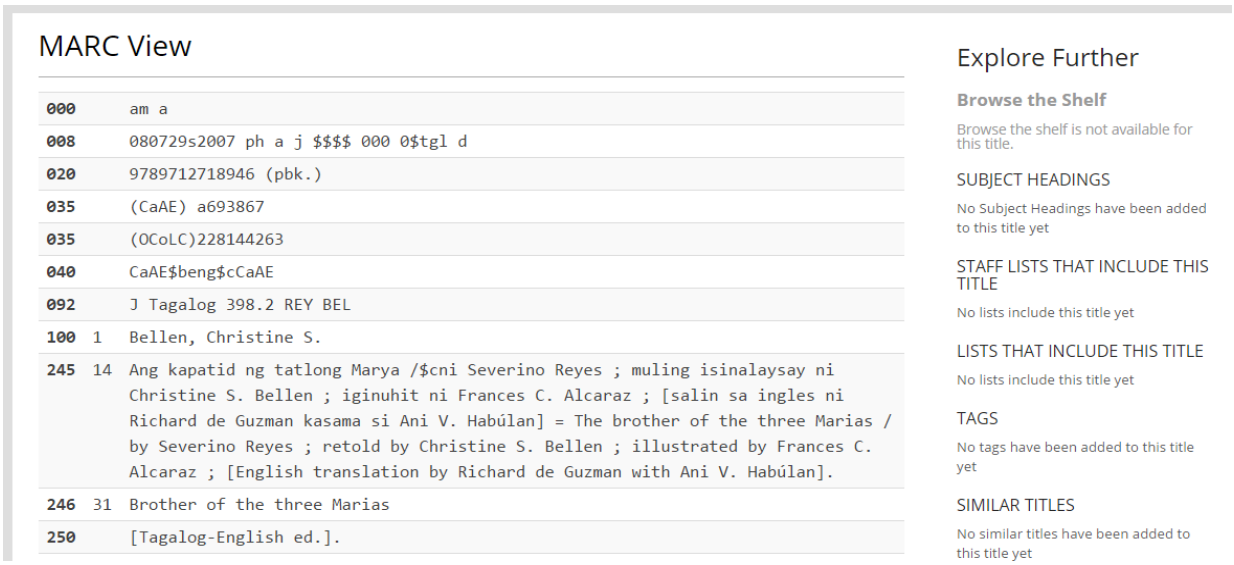


Figure 4. A MARC record found in L2's SERP

Some of L9's results might also be confusing to searchers. The snippets displayed in some of L9's SERP include user reviews in a potentially confusing format. One of the benefits of L9's discovery layer, BiblioCommons, is in the inclusion of valuable user-generated content such as user reviews of books. However, when these reviews are chosen as a snippet text for a result, confusion may ensue. In the result shown in Figure 5, the title identifies the book in question, while the snippet text appears to be a negative review of the book. Having this content as the only description of the page a searcher sees in results may cause this link to appear unprofessional, or make the searcher unsure whether authoritative information about the title is offered by the page. It is not always possible to control the content of rich snippets, because snippets are generated in response to specific queries. However, some control can be imposed by indicating which sections of a page are meant to serve as a general description using appropriate markup.

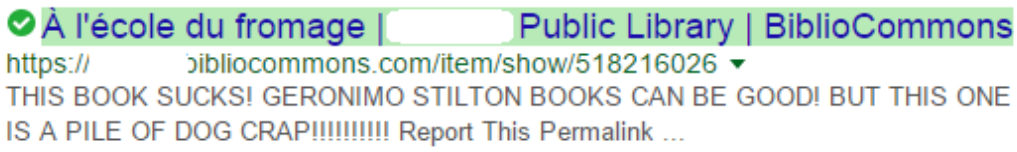


Figure 5. A user review found in L9's SERP

4.5.4 The Library with the Most Detailed SERP

L6 has its own system for displaying catalogue content online, rather than using a vendor. Its catalogue is an integrated part of its main website domain, making a comparison of main site versus OPAC impossible. Search result titles usually include the library's name, as well as a specific identification of content (a book's title, or the name of a service). Snippets are also quite specific and tend to include useful information (e.g., "This well-loved poetry collection contains outrageously funny poems and drawings"). Results include everything from library hours and library programming, to book lists and specific resources from the catalogue.

L6 has by far the highest number of results in site-specific searches: over six million pages are indexed, including all different types of library services and resources. This high number of results is not necessarily a positive thing, however. According to P6, L6 receives warnings from Google Webmaster Tools about having too many URLs (see Section 4.3.3.2). This is because L6 uses faceted search to help users navigate its catalogue, which results in many different URLs being created for the each resource. Because L6's system does not specify a canonical URL for each resource, search engine crawlers index every repeated version. Because of this proliferation of URLs, although the results retrieved in my site-specific searches are of a high quality, and include all of the right resources, library resources may not be rising as high in SERP as they otherwise would.

4.5.5 Libhub Early Adopters

L2 and L8 are both part of the Libhub early adopter program. This means that both libraries have another domain, provided by Libhub, which is intended to make library resources visible in SERP. Even within Libhub-provided domains, however, the results that appear in a Google search vary.

L2's Libhub domain has about 71,900 search results in Google. Some results include the title "Library.Link Network" followed by the title of the resource in question. Others

include the name of the library itself. Each result also includes a snippet of information, however in some cases this information is less intelligible than others. Some snippets contain details about the resource in question, but the majority contain what looks like URLs used to classify the resource:

> Type. http://bibfra.me/vocab/lite/Work; http://bibfra.me/vocab/marc/Audio; http://bibfra.me/vocab/marc/Musical; http://bibfra.me/vocab/marc/Music. Label: Concerts ...

This type of snippet may not be attractive to searchers.

On the other hand, L8's records, as published through the Libhub Initiative, produce about 211,000 results in Google. The titles include specific book titles and the name of the library, and each result has a snippet that adequately describes the book (e.g., "The faction-based society that Tris Prior once believed in is shattered, fractured by violence and power struggles and scarred by loss and betrayal. So when ..."). It is unclear why there is such a difference in the search results between these two domains, both of which are provided by Libhub.

4.6 WEBSITE TRAFFIC DATA

4.6.1 Website Traffic Comparisons

As noted in Section 3.6.2.2, libraries provided differing amounts of website traffic data, depending on each library's data collection capability and comfort-level with sharing data. Five of the ten participating libraries provided Google Analytics data on the source of hits for their main websites, while six of the ten provided Google Analytics data on the source of hits for their OPACs.

L6 also provided data but is not included in those counts because it does not have a separate OPAC domain. Without that split in traffic between a main site and an OPAC domain, L6's results cannot be compared with other libraries' main site or OPAC statistics, and must be considered separately. However, the difference between L6's data

and other libraries' data is interesting in and of itself, as will be shown in Section 4.6.2. L4's statistics are also somewhat different, because L4 is a consortium that manages many different libraries' main sites, as well as one central OPAC and a central discovery layer. Because L4 does have a separate OPAC domain, it is possible to draw on that data, as well as include main website data from a representative system within the consortium. For this reason, L4 is included in the comparative charts below, but its fundamental differences must be borne in mind. Even for the rest of the libraries that provided data, the results are not strictly comparable, because they include data from different periods. However, a broad overall comparison reveals some interesting patterns.

A comparison of the available data reveals an overarching pattern in how users find library resources online. Participants in this study described their own ideas of how users are reaching their libraries' resources, and in this case the quantitative data collected supports their impressions. For the libraries with separate main site and OPAC domains, the consensus seems to be that users navigate to libraries' main sites first, then follow some kind of link or portal to the resources contained in their OPACs:

> People, in our case anyway, they have to come to our webpage, then they have to find a link to take them to our catalogue, *then* they have to search the catalogue. (P10)

> We strongly believe that the most [common] way that people find it is they come to www.xpl.ca and they search, they enter their first search in the search box that we have at the top of our webpage, and that's how they find the catalogue. (P7)

P9 sums it up succinctly: "the library website is really a referral mechanism to the catalogue" (P9). The data reported below supports this conclusion.

Figure 6 compares the three most popular sources of traffic received by each library's main website: direct hits, organic search, and referrals. "Direct hits" refers to traffic that reaches the website directly: users have either bookmarked the site's URL, or typed the URL into their browser manually in order to reach the site. "Organic search" refers to

traffic that reaches the website through search engines such as Google and Bing. "Referrals" refers to traffic that reaches the website by following a link from another site (for example, a user may be referred from a municipal government website to the library's site). Other sources tracked by Google Analytics include hits reaching the site through email clients, and hits reaching the site through social media; however, these sources are omitted from the figures below because they each account for less than 1% of hits for each library. For this reason, the percentages shown will not add up to 100%. The missing percentage points correspond with the omitted data from email clients and social media. While the specific percentages differ between libraries, a clear overall pattern emerges: direct hits are by far the most popular path to libraries' main websites.

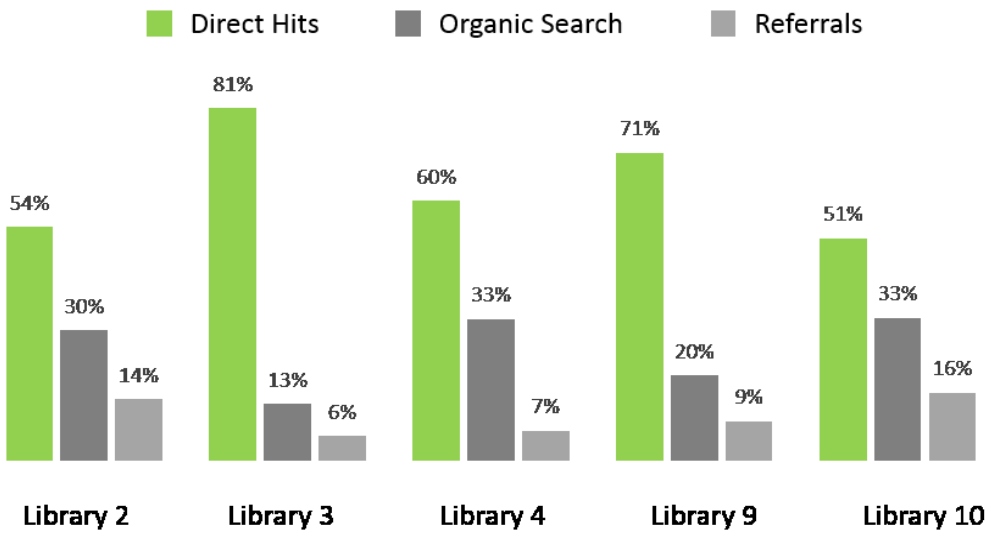


Figure 6. Percentage of Traffic to Main Sites by Source

Figure 7 compares the three most popular sources of traffic received by each library's OPAC: Referrals, direct hits, and organic search. Again, for each library, the percentages shown will not add up to 100%, because data from email clients and social media is not shown. Referrals are by far the most popular source of hits, with direct hits coming a distant second, and organic search barely registering on the scale.

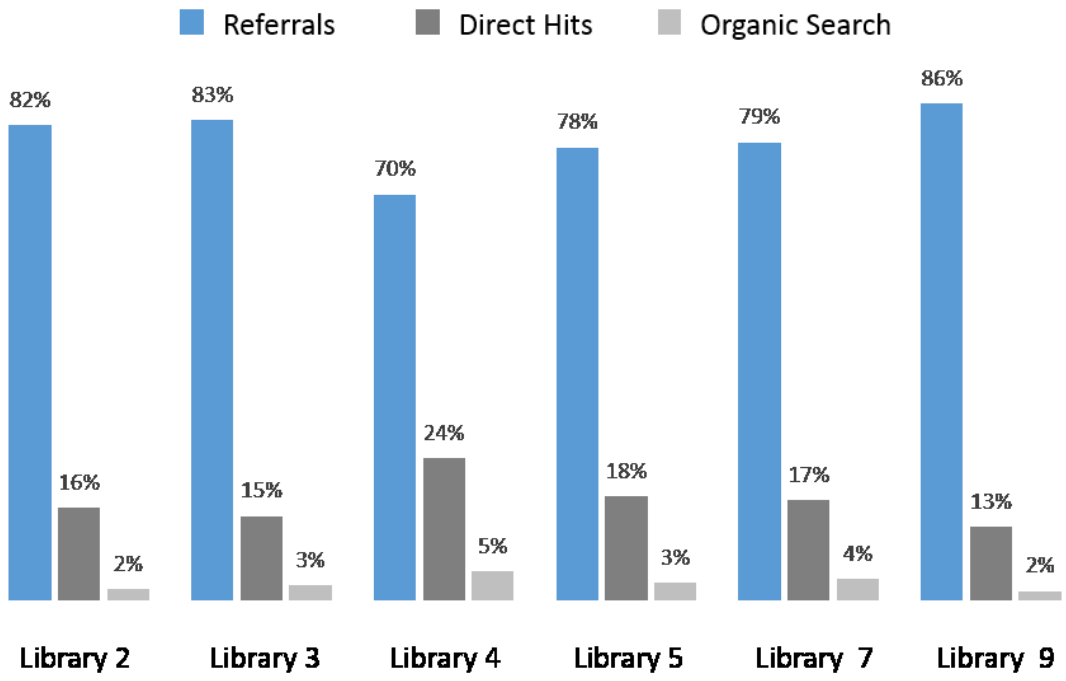


Figure 7. Percentage of Traffic to OPACs by Source

Referrals account for a minimum of 70% of hits, but from where do those referrals originate? A look at the top referrers for L2's OPAC shows that 97% of referrals come from L2's main website. When taken together, this data implies that a typical user's path to library resources is to navigate directly to a library's main website, and follow a link to the library's OPAC in order to search for resources.

In most cases, the website traffic data provided by participating libraries was not in sufficient detail to warrant a separate analysis. However, L6, as the only library with organic search as its most popular avenue of discovery, and L2, as the only Libhub participant to provide data, both warrant a closer look. The following sections (Section 4.6.2 and Section 4.6.3) investigate the data provided by these two libraries in more detail.

4.6.2 Library 6

As the only library included in this study with one unified domain, rather than separate main site and OPAC domains, L6 does not fit in with the other libraries' patterns of

traffic. As shown in Figure 8, L6 is the only library to have organic search as its most popular source of traffic.

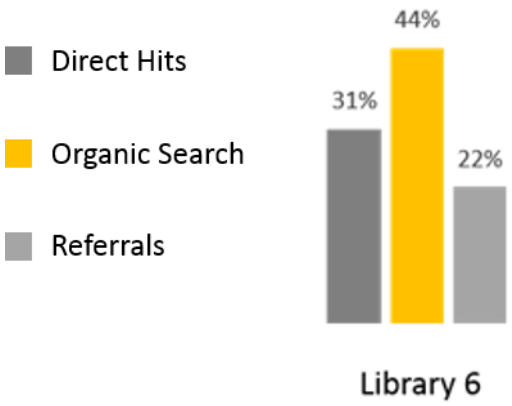


Figure 8. Percentage of Traffic to L6's Website by Source

Search engines such as Google and Bing are the source of 44% of L6's traffic. This suggests that L6 has comparatively strong visibility in SERP. This fits with P6's impressions concerning how users find her library: "Google is massive, it's a huge, huge source of the traffic to our website" (P6).

L6 also shared data on which types of search engine queries bring users to the library's website. This data can help illuminate from where that 44% of organic search traffic originates. Table 5 shows an anonymized dataset containing the top thousand queries used to reach L6's website over a three-month period. During analysis, each of the top thousand queries was categorized into six types, shown below. Queries including the library's name (e.g., "x public library" or "xpl") were by far the most popular, accounting for 79.73% of the top thousand queries. The second most popular type of query, at 15.71%, did not include the library's specific name, but did include library-centric keywords within the query (e.g., "library card", "library hours", and "ebook library"). Third on the list, with only 3.28% of queries, are queries about general information resources that make no specific mention of libraries (e.g., "jstor", "mango languages", "book club (city)"). Only 0.76% of the top thousand queries were searches for specific

resources without any library-related keywords. The last category, "Other", includes queries that could not be interpreted clearly enough to categorize.

Table 5: The top thousand search engine queries used to reach L6's website, categorized by type and shown as percentages

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%

Based on the types of keywords included, some tentative inferences can be made regarding the intention of these queries. This allows the dataset to be divided into two broad categories: queries which appear to intentionally search for the library (those including the library's name or general library-related keywords), and queries that search for resources without any expressed awareness of the library. This second category, comprising queries including the names of general resources without library-related keywords attached, and queries specifying specific resources, represents that untapped audience libraries are trying to reach with visibility initiatives. Being visible in search engine queries made by current library users is important too, but the need to reach out to non-library-users is a major motivating factor for many of the participants of this study (see Section 4.2.2). Even with all the traffic being generated by search engines for L6, this category of queries makes up only 4.04% of the top thousand queries.

Again, this data is consistent with P6's understanding of L6's website traffic:

> If you go deep down you will find searches for specific book titles, that have driven traffic to our website, but they're way down in the results. Like, tiny. I dug around and saw that one of them was "Canadian Fundamentals of Nursing, 5th Edition". If it's a specific enough search, and there's not that many copies of this book or not that much information about this book

> available elsewhere... Someone did end up clicking through from a Google search for that specific book to our website, but that was like one person. (P6)

The specific resources included in this set of top queries do tend to follow the pattern P6 describes. The few specific resources that made it into the top thousand queries appear to be titles that might have less online information available, which may account for why L6's resources made it to the first page of users' SERP.

Providing specific examples of queries for individual resources risks identifying L6, because a reader entering those queries into a search engine might see L6's name in the list of results. However, an overview of the types of materials included in those queries is illuminating. Queries include searches for textbooks, vocabulary aids, language-learning resources, relatively obscure fiction and nonfiction titles, and books in languages other than English. The most popular language appears to be Tagalog. For instance, one particular children's fiction book written in Tagalog accounts for 937 clicks when one combines various different versions of the same query (i.e. different spellings and versions of the same title). Generally speaking, the volume of traffic caused by these individual resources ranges from approximately 40 to 200 clicks each. The number of clicks being generated by these specific-resource-related queries, mostly fewer than 100 clicks per query, are relatively low, especially in comparison to the most popular query. L6's name, spelled correctly, accounts for 256,009 clicks over a three-month period.

4.6.3 Library 2

Library 2, as the only Libhub participant to provide website traffic data, also merits closer examination. In the time between my conversation with P2a in 2015, and my conversation with P2b in 2016, L2 has moved from the initial planning stages to implementation of a visibility initiative. P2b informed me that as an early adopter of the Libhub Initiative, L2 has converted a significant amount of bibliographic records into the BIBFRAME format, and published those records through Libhub. These records are fully indexed by search engines; if a searcher clicks on a Libhub link, they will be redirected to appropriate page in L2's OPAC.

Because P2a and P2b each provided website traffic data for L2 from different periods, a comparative analysis is possible. P2a provided data for the entire year of 2014, while P2b provided data for the entire year of 2015. Figure 9 shows a comparison of the traffic to L2's OPAC for each year.

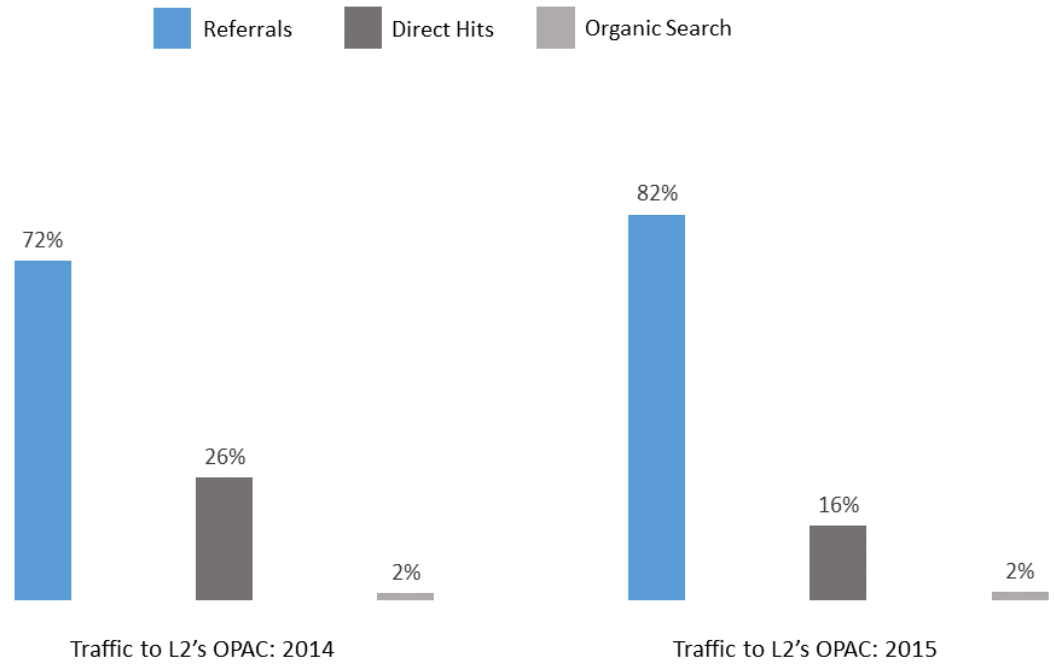


Figure 9. A comparison of traffic to L2's OPAC from 2014 to 2015

While the percentage of users reaching L2's OPAC through organic search has not changed, the percentage of users reaching L2's OPAC through referrals has increased. An increase in referrals could potentially reflect the impact of the Libhub Initiative; although Libhub reaches out using search engines, technically speaking users enter the Libhub page via organic search, and then are referred to the library's OPAC. This would cause the Libhub Initiative to contribute to the library's overall referrals, rather than its organic search numbers. Looking at a comparison of the sources of referral traffic helps to illuminate this data.

Table 6. Top sources of referrals to L2's OPAC in 2014

Top Sources of Referrals in 2014	
Source of referrals	**Percentage of hits**
L2's Main Site	97.22%
Test page	0.52%
worldcat.org	0.35%
Unknown IP address	0.18%
Interlibrary Loan Page	0.16%
Interlibrary Loan Worldcat page	0.15%
Email 1	0.09%
Email 2	0.09%
Email 3	0.07%

Table 7. Top sources of referrals to L2's OPAC in 2015

Top Sources of Referrals in 2015	
Source of referrals	**Percentage of hits**
L2's old main site	95.92%
L2's OPAC site	0.68%
L2's new main site	0.38%
worldcat.org	0.38%
Interlibrary Loan Worldcat page	0.31%
L2's summer reading club page	0.27%
Avatar creation page	0.26%
L2's new Libhub site	0.19%
Interlibrary Loan Page	0.17%
Unknown IP address	0.14%

Tables 6 and 7 reflect the changes L2 has undergone over the last two years. Not only has L2 become a Libhub early adopter, L2 also has a new main website. Although L2's Libhub site does show up on the list of referrers, it only accounts for 0.19% of referrals in 2015. This data is in contrast to P2b's perception of Libhub's impact:

> It's difficult to say, because I haven't done a comparison of pre- and post-involvement in the initiative. But seeing that the Libhub sites are 3rd and 4th

> in the top sources for referral traffic, I think that we have increased the visibility of our collection online. (P2b)

One possible explanation for this discrepancy is that, as P2b informed me, she was looking at data from only the previous month when making the above statement. It is possible that L2's Libhub sites (an older initial version and a new, recently launched version) do indeed account for a large portion of referral traffic over recent months. However, the data P2b shared represents the entire year of 2015, and during a portion of that year, the Libhub sites were not in operation. This would cause them to constitute a relatively low percentage of total referral traffic, because there were months when all the other potential sources were contributing hits and they were not. It was not possible to obtain data for a smaller period; this causes my analysis to be somewhat inconclusive. A more exhaustive exploration of L2's website traffic data would require more complete access to that data.

It is still fairly early in L2's involvement as a Libhub early adopter, and P2b has yet to investigate the associated data in detail. However, P2b identifies performing a detailed analysis of website traffic data, and evaluating Libhub's impact, as the next step for her library's visibility initiative:

> The next step is to really analyze the data using Google Analytics and assessing pre- and post- involvement in terms of referral traffic, and also seeing […] if we can follow a user's path, so once they come from Libhub, land on our catalogue, see what activities we can analyze from there. (P2b)

4.7 Summary of Results

This chapter has reported the data collected during this study, describing a complex array of factors that may motivate, inhibit, or enable search engine visibility in Canadian public libraries. Three motivating factors were reported: the need to adapt to changing search behaviour, the desire to reach out to non-library-users through search engines, and the possibility of contributing to the overall quality if information available to the public through search engines. Unsurprisingly, given that most participating libraries have yet to

achieve search engine visibility for their resources, barriers were the most common factors to emerge in this study's results. Barriers were classified into five categories: attitudes, organizational issues, technical difficulties, limited resources, and vendors. The enabling factors identified in this study also fell into five categories: inter-library cooperation, attitudes, organizational enablers, technological enablers, and vendors. Data gathered through site-specific searches of library catalogue domains (Section 4.5), as well as website traffic data provided by the libraries themselves (Section 4.6), was provided as a triangulation point for the analysis of the factors identified in the interviews. While the factors reported here may not be present in every Canadian public library, the fact that they are reported by some of this study's participants is significant, and provides a foundation for further research on the topic.

CHAPTER 5 – DISCUSSION

The motivating, inhibiting, and enabling factors revealed in this study are reported separately in the Results chapter, but these factors are part of a complex and interacting system. In Sections 5.1, 5.2, and 5.3, these factors are discussed in relation to the three research questions stated in Section 1.1. With so many interrelated factors involved, it can be difficult to decide where energies should be focused to overcome barriers to visibility. Based on my analysis, I suggest that addressing negative attitudes about search engine visibility, and clarifying and mitigating technical difficulties, are two productive ways to address this complex problem. Section 5.2.1 contains an analysis of the attitudes reported in Chapter 4, with a discussion of some possible explanations for those attitudes. In Section 5.2.1.1, an alternative perspective is discussed to help offset negative attitudes, and support the importance of search engine visibility. Section 5.2.3 outlines the participating libraries' current level of visibility, examines the technological difficulties reported in Chapter 4, and offers possible solutions to each barrier.

Beyond the library-focused factors discussed here, search engine companies are important and unpredictable factors. These companies' decisions regarding relevance algorithms, SERP display, the Semantic Web, and so forth, will inevitably have an impact. Speculation regarding search engines' motivations or intentions is beyond the scope of this study, as no interviews were performed with search engine representatives. However, publicly available information about how search engines work, as well as the SEO and metadata best practices suggested by search engines, does inform my discussion of technological barriers and enablers.

5.1 RQ1: What Factors Motivate Libraries to Address Search Engine Visibility Issues?

The data reported in Section 4.2 addresses Research Question 1, revealing three different factors that may be motivating Canadian public libraries to prioritize search engine

visibility for their resources. Of the twelve participants in this study, eleven express at least some awareness of search engines as outreach tools with the potential to make library resources more accessible. Participants identify the public's changing search behaviour, and a need for libraries to adapt (see Section 4.2.1). While participants do not doubt their libraries' relevance to their specific communities, fears about libraries' perceived relevance in the Google era are very much still alive. Increasing libraries' visibility on search engines is perceived as one way of adapting to changing times and remaining relevant. Search engine visibility is also seen as a way to reach out to members of the community who might not already be library users (see Section 4.2.2). Being open and inviting to all community members is an important value for libraries, and participants identified being more visible on search engines as a step towards achieving that goal. Based on participant responses, it seems that the overwhelming popularity of search engines, combined with the possibility that members of the public may not be aware that library resources are not being included in SERP, makes the need to adapt more urgent. Not only are opportunities being lost to reach out to potential library users; if people are unaware of why library resources are not visible in SERP, that invisibility may erode existing relationships with patrons.

Perhaps the most interesting motivation reported in Section 4.2, however, is participants' desire to contribute to the quality of information available through search engines (see Section 4.2.3). As with the other motivating factors discussed, this motivation aligns well with libraries' traditional role as information providers. Librarians are skilled in assessing, curating, and providing high quality information; search engines also assess the relevance of information and direct users to information resources (Thurow, 2015). While at times this similarity in mandate has caused search engines to be viewed as competition (see Section 2.1.1), it also suggests the possibility of collaboration. This implies that while search engines have much to offer libraries in terms of outreach and visibility, libraries also have much to offer search engines in terms of authoritative data and quality information.

As librarians begin to see search engines as tools not only for personal information seeking but also for reaching out to customers and for providing searchers with high quality information, they may become more motivated to make search engine visibility a priority. These motivating factors may push libraries to develop organizational enablers such as a culture of innovation and willingness to take risks, and mitigate organizational barriers by increasing buy-in for visibility initiatives and convincing library leaders to make the issue a priority. This in turn may help libraries push past other barriers, such as a lack of resources.

5.2 RQ2: What Barriers Stand Between Canadian Public Library Information Resources and Search Engine Visibility?

This study has revealed a complex combination of perceived barriers that may be reducing libraries' ability to make their resources more visible in SERP. Section 4.3 addresses Research Question 2 by reporting data about barriers to visibility in the form of attitudes, organizational factors, technological difficulties, limited resources, and library-vendor relationships. The implications of these barriers are discussed in this section. Barriers in the form of attitudes and technological difficulties are identified as two possible starting points in mitigating barriers to visibility, and thus are discussed in particular detail. Finally, relationships between the barriers are discussed in Section 5.2.6.

5.2.1 Attitudes as a Barrier

Addressing negative attitudes may be an effective starting point in mitigating barriers to visibility, because librarians' attitudes will inevitably colour their approach to other factors. Dispelling negative attitudes may help librarians deal with organizational and technological issues more effectively. Participants in this study expressed concerns about search engines and SEO initiatives. These concerns should not be discounted, but should be weighed against the reality of search engines' influence on public perception and awareness.

Negative perceptions of search engines' commerciality cast them as an adversary, with interests in opposition to libraries' goals as public institutions. Although participants accept the necessity of a commercial motivation in undertaking a task like indexing the Internet, their underlying distrust of that motivation remains. In some cases, this distrust may discourage librarians from undertaking visibility initiatives at all. It may also reduce the effectiveness of visibility initiatives by making it more difficult to gain buy-in throughout organizations. L8, which is in the midst of an ambitious visibility initiative, is an example of this problem: lack of staff buy-in is still a challenge, even though the library's leadership has decided to prioritize the issue (see Section 4.3.2.3). While these negative feelings may be justified in some ways, they are not conducive to a productive relationship with search engines. Gaining a deeper understanding of the mechanics of indexing, relevance ranking, and so forth, might help librarians move beyond their misgivings and become more adept at navigating the world of search engines.

When this distrust is added to concerns about data security and patron privacy, librarians may feel that undertaking even basic steps towards visibility, such as allowing crawlers to index bibliographic records, could violate one of the libraries' core mandates. Privacy advocate Cavoukian (2011) argues that user privacy should be built into library systems and practices by default, with the aim of preventing privacy issues before they occur. This Privacy by Design approach requires librarians to carefully consider the potential privacy and data security implications of any technology used (Cavoukian, 2011). Some, such as Fortier and Burkell (2015), feel that even the collection of non-personally identifying information about library patrons is problematic and should be avoided.

These are important concerns for libraries, but should not necessarily stand in the way of visibility initiatives. The general lack of comfort with, and understanding of, technology identified by participants within their organizations (see Section 4.3.2.2) may be magnifying security concerns, at the same time as it reduces librarians' ability to address those concerns in constructive ways. It is difficult to assess the potential risk of such technologies without a clear understanding of how they work. Gaining a clearer understanding of the technology involved may allay some of these concerns, and bolster

librarians' confidence that they are taking all the appropriate steps to increase their resources' visibility without compromising their patrons' privacy. Conversations with vendors of library catalogue technology may also be necessary, in order to ensure that patrons' private data is being stored in such a way that allowing crawlers to index the catalogue domain will not compromise patron privacy.

Concerns about gathering data using tools like Google Analytics could also be addressed by gaining clarity about the data collected, Google's privacy and security policies, and the exact nature of any potential privacy risks. A first step for librarians wishing to improve their understanding of the privacy implications of Google Analytics is a careful review of Google's general privacy policy[1], Google Analytics security and privacy principles[2], and the Google Analytics terms of service[3]. These materials make it clear that "sending personally identifiable information (PII) to Google Analytics" is prohibited (Google, 2016d, para. 3), and require all Google Analytics users to have an appropriate privacy policy that includes an explanation of the data gathered using the tool (Google, 2015). They also outline methods by which libraries can offer their patrons the opportunity to opt out of Google Analytics data collection (Google, 2016d). These materials may allay privacy concerns by showing that at least some of those concerns are shared by Google, and also by offering practical resources for helping library patrons protect their privacy. This issue still requires careful thought on the part of librarians, depending on their views about the collection of behavioural data in general. Some librarians may still conclude that collecting any data about patron behaviour on their websites is problematic. However, for those who are willing to engage in this type of analytics, Google offers a certain amount of support to aid in understanding the privacy implications of tools it offers, if librarians can set aside their distrust and make use of such resources.

---

[1] Available here: https://www.google.com/policies/privacy/
[2] Available here: https://support.google.com/analytics/answer/6004245
[3] Available here: https://www.google.ca/analytics/terms/us.html

The perception that true search engine visibility is an impossible goal may also slow adoption of visibility initiatives. The technological challenges involved with visibility are real, and in some cases have no simple resolution. A serious effort to become visible through search engines would be time-consuming and costly; and, as P7 says, "perhaps it would fail". While participants feel that libraries should be "invading that space" (P4), they also worry that making the attempt would be an act of hubris akin to librarians' desire to "catalogue the Internet" (P7) in the early 1990s. Will search engine visibility be the next big revolution in community outreach, or will it be an expensive and embarrassing misstep? There is no way to be certain. This possibility may cause libraries to hold back, waiting for others to succeed before taking on the challenge.

Concerns about the viability of search engine visibility for libraries, as well as concerns about the profit-centred motivations of search engines, are reinforced by past experience. Librarians have been burned before when it comes to the internet, and search engines specifically. West's (2015) chronicle of libraries' failed love affair with Google is a playful portrayal of events that left a very real impression on librarians. Google's apparent interest in libraries during the 2000s, as shown by the company's presence at American Library Association conferences and the creation of Google's (now defunct) Librarian Central blog, had faded to apparent indifference by the early 2010s (West, 2015). West (2015) suggests that librarians accepted Google's initial interest in good faith, and feel betrayed by Google's defection. Although West's article is a personal opinion piece and makes no claims to speak for all librarians, Section 4.3.1.1 seems to confirm her depiction of librarians' feelings, at least among these study participants. Google's indifference is still keenly felt: search engines "don't really care about libraries" (P1a), and they may not have "best interests of the public at heart" (P9). Taking a chance on future cooperation with search engines requires librarians to move beyond those negative memories and misgivings, with no guarantee of success, and without any particular encouragement from search engines. The most straightforward solution to this problem would be for some libraries to take the plunge, and through their experiences provide more data about whether true visibility is possible for libraries.

Even if libraries are successful in becoming more accessible through search engines, the possible outcomes are not seen as purely positive by all participants. As shown in Section 4.3.1.3, some participants believe that a successful visibility initiative would have unknown and possibly far-reaching consequences. As a major technological innovation, search engine visibility has the potential to alter the current dynamic between vendors and libraries, with unpredictable outcomes. In addition, if the result of increased visibility is increased use of library systems, that in itself may be cause for alarm. Increased demand for library services is positive, in that it confirms libraries' relevance and strengthens libraries' claim to funding. However, if demand exceeds capacity before funding is increased, libraries may have difficulty keeping up. Library systems with limited resources may be daunted by the idea of opening up their resources to what amounts to a global audience, and potentially receiving demands for their services that are beyond their ability to fulfill.

Participants' mixed feelings about increasing the use of library services are matched by mixed feelings about increasing libraries' profile online. While most participants are not actively opposed to becoming more visible, P3, P9, and P1b note that the issue is not seen as requiring any particular urgency in their organizations. Their libraries are already quite successful in reaching members of their communities, and thus are perceived as not needing added exposure. As P1b notes, this attitude is in stark contrast to the approach taken by most commercial companies. Libraries, as prominent public institutions, seem to be satisfied with "enough" patrons, and "enough" exposure. There is a certain level of confidence that enough community members will always use and appreciate the library to support the institution's perceived value in the community. Once it successfully reaches a high percentage of its target population, a commercial company would not necessarily see this as a reason to stop trying to reach more customers. However, as Section 4.3.1.1 shows, libraries are decidedly, and purposefully, non-commercial entities. Osif (2006) notes that this purposeful non-commerciality can reduce librarians' willingness to engage in any form of marketing: "For many the very thought of selling might result in a shudder or cringe. There is a purity, a higher purpose to librarianship and a somewhat less high, less pure image to that of selling something" (p. 39). Perhaps this attitude rightly reflects

libraries' priorities, which centre on providing services rather than promoting services. However, waiting for a crisis in engagement before adapting to changing search behaviour may be problematic in the long term. The deeper question here, which this study does not pretend to answer, is whether libraries should be in the business of actively promoting themselves.

5.2.1.1 Evidence to Support the Importance of Search Engine Visibility

One way to mitigate the effects of negative attitudes about search engine visibility is to introduce new information to the conversation. Without discounting participants' concerns about search engines, this section provides some evidence to place on the other end of the scales. Search engines may be commercial. They may not be considering libraries' needs. However, they also present libraries with an important opportunity to connect with the public. Libraries may be well-established and important institutions, but in a time when how people access information is changing so rapidly, relying on that established status may be a mistake.

Whether or not libraries should be in engaged in self-promotion, Figure 10 (below) challenges the perception that libraries do not need added exposure. Figure 10 is based on Google Trend data on searches made about the books as a general topic and libraries as a general topic from 2004 to 2015 (Google, 2016a). Google calculates the relative popularity of each topic by comparing the total number of Google searches related to that topic with the overall number of searches performed in Google over time (Google, 2016b). Figure 10 shows a gap in popularity between the two topics. Even more telling, the popularity of books as a topic has remained relatively stable, while the popularity of libraries as a search topic has declined slightly over the years.

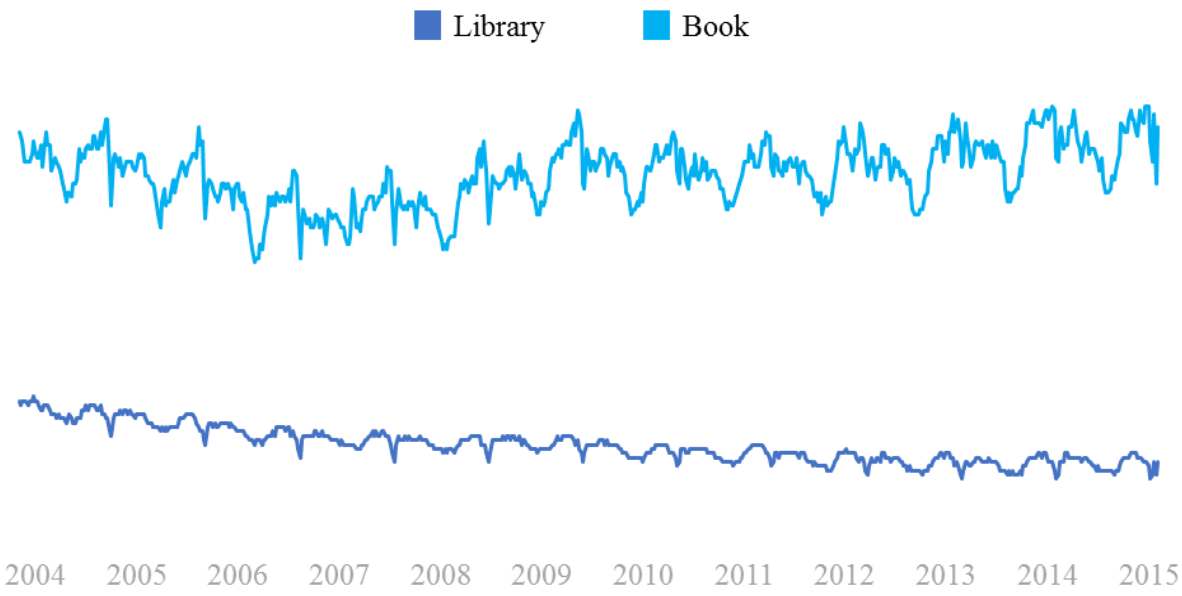


Figure 10. Google search trend data on the relative popularity of library-related and book-related queries from 2004-2015

It should be noted that this chart is only partially applicable to the question of Canadian libraries' visibility. Libraries offer much more than books, and this chart shows worldwide data, rather than Canadian-specific information. In addition, not every searcher entering queries involving books is actually interested in reading the book in question, nor would every searcher be interested in acquiring books from the library if presented with the opportunity. Nevertheless, it is reasonable to suppose that a large number of those making book-related queries might be interested in the many services offered by their local library. Based on the comparatively low number of library-related queries, it is also reasonable to conjecture that a significant number of these searchers may not consider libraries as an option. If nothing else, the gap between the search volume for books and the search volume for libraries shows that search engine users would be fertile ground for library outreach. Google searchers appear to include a population with a high interest in one of the core services libraries provide, but a comparatively low awareness of libraries as institutions; this population is a perfect target for outreach.

Another important consideration in deciding whether the risks of prioritizing search engine visibility are worth the potential gains is the question of libraries' perceived relevance as public institutions. Canadian libraries do essential work in their communities, and make many important contributions to the wellbeing of community members. However, that work will continue to be effective only while people continue to see libraries as relevant to their needs. While some participants have no doubt about their communities' awareness of, and engagement with, library services, it has been shown that search engines have an increasing impact on perceived relevance in both the online and offline domains. If libraries do not become more visible on search engines, this impact could eventually be felt in public perceptions of libraries' relevance.

Users place a high level of trust in Google's relevance ranking. For instance, Balatsoukas and Ruthven (2012) used eye-tracking technology to examine the decisions about relevance people make when using Google. Their results revealed "users' blind trust in Google's ranking algorithm and specifically their tendency to click, without any critical thought, on the first surrogate presented in the result list" (Balatsoukas & Ruthven, 2012, p. 1736). Search engine results can influence not only the volume of traffic a site receives, but also general awareness of a product or organization. Dou et al.'s (2010) study tests the theory that search results can impact awareness of brands, "regardless of whether people click on them" (p. 262). Because "internet users expect the most relevant sources to be listed at the top of a SERP", Dou et al. hypothesized that an unknown brand's position in the list of results would impact the likelihood that a searcher would remember the brand later (p. 264). Their experiments showed that when an unknown brand appeared above known brands in search results, subjects were more likely to remember the unknown brand when asked about it in a later recognition test (Dou, Lim, Su, Zhou, & Cui, 2010). This research shows the impact of SERP not only on website traffic, but on overall awareness of brands, implying that increased search engine visibility for libraries could not only increase traffic to library websites, but also increase public awareness of libraries as a relevant community service.

The arguments presented in this section are not meant to entirely erase negative attitudes about search engine visibility, but simply to add an alternative and productive perspective. As P2a says, "As a public librarian […] you want to be where the people are." Multiple studies have shown that people use search engines for their information seeking needs (DeRosa et al., 2011; Purcell, Brenner & Rainie, 2012). In order to fulfill this mandate and support their communities' information needs, librarians must move out of their comfort zones, move beyond their misgivings, and engage more fully in this extremely popular information forum.

5.2.2 Organizational Barriers

The organizational barriers reported in Section 4.3.2 may stop search engine visibility from rising to the top of priorities, or being seen as a goal at all. The data collected in this study supports Arlitsch, O'Brien, and Rossman's (2013) proposal that library administration must become more involved with SEO: if search engine visibility is not being considered or understood at the highest levels of organizations, where institutional priorities are set, the issue will likely not be addressed. Lack of familiarity or comfort with the technology used for search engine visibility may also have an impact, making it difficult to obtain buy-in for search engine visibility at every level of the organization. Because search engine visibility often falls under the purview of multiple departments, but is seldom included in the explicit tasks assigned to any of those departments, it is difficult for organizations to address the issue or even be aware that it exists. Some libraries, such as L2, have mitigated this problem by creating a multi-departmental taskforce to address this type of visibility. This approach may be effective, as it builds an official structure for dealing with search engine visibility. However, as yet, few organizations involved in this study have taken that approach.

5.2.3 Technological Factors

Addressing technical difficulties experienced when attempting to increase the visibility of library resources may also be a productive place to begin mitigating barriers to visibility. Because some technical difficulties have comparatively simple solutions, overcoming

these barriers can make an immediate positive impact on search engine visibility in return for a relatively low investment of time and effort. This section begins by describing the context for these technological factors through an analysis of the current visibility of participating libraries, and a discussion of what successful visibility for library resources means in this context. Section 5.2.3.1 discusses the search engine visibility of participating libraries based on the data reported in Sections 4.5 and 4.6. The technological barriers perceived by participants are then addressed in Section 5.2.3.2, and possible solutions are suggested.

5.2.3.1 Libraries' Current Level of Visibility

The website traffic data provided by some of the libraries in this study tells a clear story: the majority of visitors reach library websites through direct access, and from there follow a link to the library's OPAC and the resources contained therein. The fact that the majority of library patrons find the library's main website through direct navigation implies that the majority of visitors are not only already aware of the library as a resource, but presumably either have the library's website bookmarked, have memorized the URL, or use it frequently enough to have their browser auto-fill the URL. While it is heartening that these libraries have so many devoted patrons, this usage pattern reveals that library websites are, in a sense, preaching to the choir. They act as an important access point to library resources, but are less successful in performing outreach to community members who are not already aware of those resources.

This is not to say that library websites perform no outreach function at all; the second most popular path users follow to library websites is organic search. However, an examination of the types of search engine queries used shows that the majority of those users were also already aware of the library before beginning their search. For instance, L6 was the only library to have organic search as the most popular channel (44%). However, an examination of the 1000 most popular queries used to reach the library website shows that the majority of queries include the library's name, or some form of

library-related keyword. This implies that the searcher was looking for a library, and often Library 6 specifically.

This is the current state of visibility for the libraries that provided website traffic data for this study. In order to even speculate about the potential of search engine visibility for libraries, one must consider what its meaning is in this context. If we take true visibility for library resources to mean that a book offered by a local public library will show up in the first page of search engine results when a searcher enters its title, without specific reference to any library-related keywords, this goal may indeed be elusive. This possible definition is based on participant comments reported in Section 4.3.1.2, and unsurprisingly leads to skepticism about search engine visibility as an attainable goal. Depending on the popularity of a certain resource, the online discourse on the topic of that resource at the time of the search, and a wide variety of other factors, a library's copy may never rise to the first SERP for a given query. It is possible that this will remain the case even if libraries implement robust SEO initiatives. With this possibility in mind, is search engine visibility worth attempting, or are such initiatives too costly for too uncertain a return on investment?

Among this study's participants, L6 has the most complete and detailed presence in SERP (see Section 4.5.4). As a result, L6 provides a useful example of what search engine visibility might look like for public libraries. Assessing the impact of L6's visibility simply in terms of traffic to the library's website, and the amount of traffic generated by organic search, implies some level of success. Forty-four percent of the traffic to L6's website comes from search engines (see Section 4.6.4). This contrasts with a maximum of 33% for other libraries' main websites and a maximum of 5% for other libraries' OPACs (see Section 4.6.1). It is difficult to compare L6 with other libraries, due to its single, unified domain name. However, the fact that the majority of its website traffic comes from search engine queries, when search engine traffic was not the most popular avenue of access for the other libraries' main *or* OPAC domains, does imply that L6's visibility in search engines is having some impact.

Nevertheless, as shown in Section 4.6.4, the vast majority of queries being used to find L6's site on search engines explicitly refer to libraries in some way. The fact that most of the searchers reaching L6's site already had the library in mind is problematic, given that reaching non-library-users is one of the major motivations for search engine visibility expressed by participants (see Section 4.2.2). In Section 4.3.1.2, P6 states that it is highly unusual for her library's resources to reach the top page of SERP, even though those resources are being indexed, unless the user includes the word "library" in the search. According to P6, there is simply too much competition for top listing when it comes to searches for individual books, and libraries' resources are "just not going to come up high enough in a general search." If this is the case for L6, which is the most successfully visible library included in this study, where does that leave libraries that have yet to begin visibility initiatives?

It is true that many search engine users are unlikely to add the word "library" to their queries; people who are not already aware of the library's services are unlikely to discover them that way. However, visibility in searches for individual resources that do not include library-related keywords is only one form of visibility. While having resources show up in searches that include "library" does not significantly aid serendipitous discovery of the library, this type of visibility may be important for different reasons. For instance, consider the possible impact of *not* having library resources show up in such a search. What if a library patron, who is aware of the library but may not be aware that its resources are not being indexed by search engines, Googles the query "book title library"? If the library's resources don't appear, will that patron be aware of the reasons for that absence, or will they simply assume that the library does not have the resource? Over time, failing to find library resources using such a query might erode confidence in the depth and breadth of the library's collection.

Although L6 is not seeing high volumes of traffic resulting from search engine queries about specific resources, a small percentage of users *are* finding the library through that pathway. In the top thousand queries reported in Section 4.6.4, queries for specific resources generally account for about 40-200 clicks each. The most popular individual

resource on the list accounted for 937 clicks. While the volume generated by these queries is low in the scale of overall traffic, it is not necessarily insignificant. If a librarian helped patrons find a specific book 937 times over a three-month period, or even 212 or 50 times, that volume of interactions might be considered significant. Especially when one considers the likely audience implied by the types of queries being used, it seems worthwhile to make a seemingly obscure resource visible to the small number of people searching for it. Searches for reading aids, language learning books, and foreign language resources are likely to be made by vulnerable populations, and populations that may not already be aware of the library. When googling a book title in their native language, a newcomer to Canada, for example, might stumble across L6's resources. That serendipitous discovery could have a powerful impact, not only by connecting the searcher with the required resource, but also by making the searcher aware of the library as a welcoming community centre with helpful resources in many different languages. This seems to be very much in accordance with libraries' mission within their communities.

Based on these considerations, perhaps the definition of success in search engine visibility should be expanded. Certainly, widespread visibility for individual resources in SERP without need of library-related keywords is an ultimate goal for visibility initiatives. However, intermediate levels of visibility, such as the current visibility enjoyed by L6, are not without value. Taking initial steps to overcome the technological barriers to visibility that are perceived by participants could provide at least that improved level of visibility. If more libraries overcome these initial challenges, their experiences will have the added benefit of providing more data about the impact of this level of visibility, and could help clarify next steps for libraries in reaching for that higher level of visibility envisioned by participants.

5.2.3.2 Technological Barriers and Potential Solutions

Some of the technological barriers perceived by participants have known solutions. Difficulties connected with older OPAC software can be solved by upgrading library

technology. A lack of links to bibliographic records can be solved with the use of traditional SEO techniques such as sitemaps, or through the linked data approach employed by the Libhub Initiative. Having too many URLs indexed due to a faceted search interface, as in the case of L6, can be addressed by defining canonical URLs for each bibliographic record, either with sitemaps, or through the use of special HTML markup on each page (See Section 1.2.2.3). These solutions are not technically difficult, but may be difficult to implement in practice if librarians do not have the ability to control vendor-provided HTML and sitemaps, or lack training in these techniques. Depending on a library's catalogue technology, these solutions may be simple to implement, or may be more difficult to put in place.

For other issues, no immediate solution presents itself. For instance, problems arising from libraries' many "silos of data" (P2a) are difficult to address because they stem from the structure of library-vendor relationships. Digital resources sourced from different vendors are difficult to integrate with library OPACs, let alone search engine indexes, and libraries are dependent for solutions on the vendors providing the resources.

Other perceived barriers may not, in fact, be barriers at all. The need for stable URLs in the indexing process, for instance, is now largely mythical. Search engine technology has evolved to make this problem a thing of the past. Google's official Webmaster Blog has confirmed that dynamic URLs no longer stop a resource from being indexed by crawlers, and will not be significantly detrimental to search engine rankings (Stiller & Szymanski, 2008). Whether or not a database uses dynamic URLs, search engines offer guidelines for ensuring that it is appropriately indexed (see Section 1.2.2). However, since many still believe dynamic URLs to be an absolute barrier to the indexing process, library staff many never attempt to implement those guidelines. Similarly, the challenge of maintaining permanent URLs, mentioned by P2a and P2b, would not necessarily be seen as a barrier in the broader online world. While maintaining permanent URLs is difficult, doing so is not generally seen as a necessary prerequisite for resources to be crawled by search engines. Content providers expect search engines to crawl their pages frequently in order to correct any URL changes that have taken place. The changeability of the

internet is widely acknowledged, and an occasional 404-error is par for the course. P2a and P2b'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.

Although allowing library resources to be indexed by search engines is known to be possible, there is still doubt about whether library resources can become truly visible in SERP. As the library in this study to have achieved the highest visibility, L6 still struggles to be visible in a meaningful way in searches for individual resources. The question of how to make library content more visible in SERP remains, to some extent, unsolved. The Libhub initiative offers one possible approach, but some participants are skeptical about its true effectiveness (See Section 4.4.5). Trying to boost library resources to the first page of users' search engine results means competing with large commercial interests such as Amazon.com, popular news outlets providing book reviews, and a slew of other content. This can be a daunting prospect. Google's (2010) *Search Engine Optimization Starter Guide* emphasizes quality of content as the most important factor in SEO, and libraries certainly have quality content to offer. This provides hope that if libraries can begin offering their bibliographic records in more search-engine-friendly ways, they will naturally begin to rise higher in SERP. However, until more library resources are optimized for search engines, there is little evidence to either support or discredit this hope.

Even if libraries succeed in having all of their resources indexed, and comply with search engines' desired SEO best practices, a lack of geolocation in book-related searches remains a serious barrier to true visibility, with no immediate remedy. Geolocation is particularly important to public libraries, compared to other memory institutions, because public libraries are so focused on local communities, and have fewer unique items in their catalogues. Aside from local history collections and the like, which constitute a small percentage of a public library's holdings, public library systems across the country offer many of the same titles. They are not unique in offering the latest James Patterson novel, but they are unique in offering access for free within their specific communities. This

similarity of titles can be a benefit for public libraries (see Section 4.4.1), but also has implications for search relevance. As a searcher in Halifax, it is less than useful for me to encounter resources from Edmonton Public Library in my search engine results. Even digital resources such as e-books are in many cases useless to somebody outside the library's community, because they require a location-specific library card to gain access. In the majority of cases, especially for public libraries with their community-centric offerings, geolocation is necessary in order to make a library resource truly relevant to a searcher.

It is unclear why search engines are apparently not putting weight on location when it comes to searches for books. The exact details of search engine relevancy algorithms are of necessity secret, to avoid deliberate manipulation (Granka, 2010). While Google shares some information about its algorithms (see Google, n.d.c), there is no explanation for why this particular type of search does not seem to be affected by geolocation. One could speculate that the prevalence of online marketplaces such as Amazon.com has caused search engines to view books as items that do not need to be sourced locally, making geolocation irrelevant. However, without more information on the topic from search engines, speculation is ultimately futile. Addressing this question in a meaningful way may require communication, and perhaps cooperation, between libraries and search engines.

5.2.4 Limited Resources

Limited resources are a reality in any organization, and particularly in those with set budgets allocated by funders. Library budgets are often stretched tight in order to fulfill all the needs of local communities, leaving little room for experimental projects. According to some of the participants in this study, allocating sufficient staff time to search engine visibility is particularly challenging. When limited resources are combined with the other barriers identified by participants, it is not surprising that only three of the ten libraries in this study have taken concrete steps toward search engine visibility.

5.2.5 Vendors

Vendors are a powerful player in library technology and innovation. This is true in many industries, but has important implications for the question of search engine visibility for library resources. As providers of library technology, vendors are seen as both barriers to, and enablers of, visibility. Libraries and vendors have developed a useful partnership that helps libraries keep their services up-to-date in an environment of rapid technological change. However, this partnership may have unintended repercussions, as libraries give up control in some areas, and risk losing touch with the technology involved in managing their content. For most of the libraries participating in this study, undertaking a visibility initiative would either involve buying a new product from a vendor (such as the Libhub Initiative from Zepheira), or would require the cooperation of a current vendor.

The library-vendor relationship raises various questions. Have libraries lost the initiative when it comes to the digital management of their information resources? If libraries become increasingly divorced from the technology they use, will librarians know what innovations to request from their vendors? A certain amount of comfort and expertise with technology is necessary to know what changes are technologically possible. As technology becomes increasingly embedded in library services, will vendors take over more and more decision-making power? Because this study did not include any participants from vendor organizations, it cannot reveal any information about vendors' motivations or priorities. However, it has revealed vendors as a crucial player in the landscape of library visibility, which suggests avenues for further research.

The library-vendor relationship is not necessarily a negative thing; perhaps it is necessary, as digital technology gets more and more complex, for libraries to give up some control in this area in order to focus attention on providing services to their communities. On the other hand, if the potential conflict of interest perceived by P3 and P9 is a reality, vendors may not be very motivated to provide solutions to search engine visibility problems. This is an area where inter-library cooperation could come into play as an enabler of visibility. As P9 points out (see Section 4.3.1.1), vendors of library

technology are ultimately motivated to help libraries remain successful, because successful libraries can continue to buy vendor technology. If enough libraries tell their vendors that making library resources visible in SERP is important to libraries' survival, vendors may become more motivated to address the issue. This may be happening already, as shown by BiblioCommons' apparent interest in the topic described by P9 (Section 4.4.5.1), and the Libhub Initiative by Zepheira (see Section 5.2.5 for further discussion of vendors as an enabler of visibility).

Participants perceive a pressing need for libraries to adapt to changing technology (see Section 4.2.1). However, the data reported in Section 4.3.5 suggests that vendors stand as an intermediary in that process. This could encourage stagnation by curbing librarians' ability to experiment on their own initiative. Regardless of the possible implications of the library-vendor relationship, it seems unlikely that libraries will suddenly pull back from vendors and begin taking back control of these technologies. Perhaps the most productive response to the potential issues raised by vendors is simply to provide librarians with more opportunities to engage with the mechanics of library technology and search engine technology, and more opportunities to learn about these rapidly-evolving technologies. This might help librarians approach the library-vendor relationship from a well-informed perspective, with a good understanding of what technological improvements can, and should, be attempted.

5.2.6 Relationships among Barriers

Barriers to visibility can reinforce each other, making it difficult to know where to begin in addressing them. For instance, negative attitudes about search engines can perpetuate organizational issues such as lack of buy-in for search engine visibility initiatives. Negative attitudes can also exacerbate technological difficulties by making library staff less interested in learning how search engines work, and thus less likely to gain the knowledge necessary to solve technological problems. At the same time, technological difficulties like those described in Section 4.3.3.3 might intensify negative attitudes by causing frustration with search engines, and reinforcing the impression that, as P1a says, search engines "don't really care about libraries". Having limited resources is an

underlying reality influencing libraries' ability to cope with all these factors. Limited budgets may force organizational leaders to prioritize maintenance over innovation, and may put the purchase of available technological solutions out of reach. This in turn can be exacerbated by organizational barriers: an issue is unlikely to rise to the top of the list of priorities if it is not well understood, falls between the cracks of departments, and has no simple or straightforward solution. Limited staff time reduces librarians' ability to look beyond day-to-day tasks at the larger strategic picture, which may perpetuate negative attitudes. Limited time may also impact librarians' ability to execute visibility initiatives with adequate support and evaluation during every stage of the project.

5.3 RQ3: What Factors Have the Potential to Enable Libraries to Achieve Search Engine Visibility for their Resources?

Research Question 3 is addressed in Section 4.4.2, where factors with the potential to enable visibility are reported in five broad categories: inter-library cooperation, attitudes, organizational enablers, and vendors. These potential enabling factors are to some extent theoretical. Since few participants would describe their libraries as having successful visibility on search engines, the discussion of enablers is in part a discussion of potential rather than actual factors. However, participants were able to speculate about factors that might enable visibility for their libraries in the future, even if those factors were not currently in place, and participants from L2, L8, and L6 were able to offer some personal experience with enabling factors encountered during their libraries' attempts to improve search engine visibility. More research is needed to ascertain the true efficacy of these factors, but identifying them as possible enablers does offer libraries possible avenues to explore in improving their resources' search engine visibility.

The enabling factors identified in this study have the potential to impact outcomes by mitigating barriers. The positive attitudes reported in Section 4.4.2 may help to counterbalance the negative attitudes reported in Section 4.3.1. P8a and P8b identified having a culture of innovation, and a willingness to take risks at the highest levels of organizations, as enabling their organization to prioritize search engine visibility despite limited resources. Technological enabling factors, such as access to accurate data about

website traffic to library catalogues, could ameliorate organizational issues by providing evidence to support the necessity of including SEO in institutional priorities. This data could also help librarians solve technological barriers by offering information about how human users and search engine robots interact with library systems, and the impact of any technological improvements. Control over the display of bibliographic records may help library staff experiment with their visibility, and develop innovative solutions to some of the barriers they face. Up-to-date OPAC and ILS technology is also an important enabler. Newer OPAC products are often more search engine friendly in their construction, with features such as stable URLs for bibliographic records that immediately address some of the perceived technological barriers outlined in Section 4.3.3.1. In addition, newer OPAC technology may free staff time by requiring less day-to-day maintenance. This may help mitigate the barrier of limited staff time and resources. Inter-library cooperation could help librarians solve technological difficulties by providing an opportunity to compare notes and share lessons learned. Communication and cooperation among libraries may also motivate vendors to prioritize SEO, if libraries make a concerted effort to discuss search engine visibility concerns with their vendors. Because search engine visibility does raise concerns for libraries, and involves a certain amount of risk, collaboration among libraries could be a key factor in achieving success.

Section 4.4.5 addresses vendors as a potential answer to RQ3. Since they can be both barriers and enablers, vendors have a large impact on the technological issues involved with search engine visibility. Participants expressed hopes that library vendors would prioritize search engine visibility, and one participant (P9) suggested that vendor BiblioCommons has already begun to do so. The partnership between libraries and vendors could be a fertile space for innovation. Zepheira's Libhub Initiative, for instance, is a promising example of a vendor taking a strong interest in the visibility of library resources. Nevertheless, there is some doubt whether Libhub is an appropriate response to that question. The concerns about Libhub reported in Section 4.4.5.2 must be considered. P6 and P9 suggest that Libhub unnecessarily conflates library cataloguing formats and linked data with the visibility of library resources. It is true that there are many important steps libraries can take to make library resources visible in search

engines that do not involve linked data. Libraries can implement sitemaps, modify their OPACs' Robots.txt files, and even add appropriate schema.org markup to records without transitioning from MARC to BIBFRAME. Perhaps P9 and P6 are correct in suggesting that adding a transitional cataloguing format into the mix is simply an unnecessary complication, which obscures the true issues involved.

Despite these concerns, the Libhub model does have potential that is still in the process of being tested. According to P8b (see Section 4.4.1), the true promise of Libhub is in the links shared among libraries. In theory, interlinking different libraries' bibliographic records should push those individual records higher and higher in SERP by combining the libraries' online reputations as calculated by search engines. This may offer an answer to the as-yet-unsolved problem of how to get library resources high enough in SERP to compete with big companies like Amazon. However, this type of interlinking must be approached carefully. Google is wary of webmasters using links to purposely manipulate relevancy algorithms, and explicitly states that "Any links intended to manipulate PageRank or a site's ranking in Google search results may be considered part of a link scheme and a violation of Google's Webmaster Guidelines" (Google, 2016e, para. 1). Unlike true link scams, the links among libraries are based in a genuine connection, but it is hard to know how this will be interpreted by search engines. While Libhub's intention is not to manipulate algorithms in any underhanded way, it is unclear at this point whether the Initiative's interlinking will be helpful in promoting library content, or be considered a link scam and actually be detrimental. It is hard to speculate concerning the outcome of the Libhub Initiative at this point, but there is at least some basis for its relevance to search engine visibility.

Some of the data collected in this study offers insight into the effectiveness of Libhub as an enabler of visibility. It is difficult to assess the success of the Libhub initiatives in L2 and L8, since they are still in the early stages of their development. Participants from L2 and L8 expressed enthusiasm about Libhub, as reported in Section 4.4.5.2, but were unable to make definitive statements about the Initiative's impact because they had not yet performed any rigorous analysis of pre- and post- involvement website traffic. As

shown in Section 4.5.5, the bibliographic records published by L2 and L8 through Libhub are being indexed in search engines. The SERP produced by these resources have varying degrees of quality in terms of their usefulness to a searcher. L8's results appear to include more useful detail, both in the titles and the rich snippets, based on the criteria highlighted in Google's *Search Engine Optimization Starter Guide* (Google, 2010). However, in both cases Libhub has clearly been successful in allowing library resources to be indexed by search engines. It is more difficult to evaluate success when it comes to the placement of these resources in SERP; are enough resources rising high enough in SERP to truly increase the visibility of each library? At this time, there is insufficient data to draw conclusions regarding this question. The only website traffic data available, shown in Section 4.6.3, is inconclusive.

5.4 Summary of Discussion

As discussed in Section 5.2.3.1, for most of the libraries participating in this study, the current search engine visibility of library resources is quite low, especially in queries that do not use library-specific keywords. The data gathered in this study provides the beginning of an explanation as to why this is the case. This chapter has discussed the motivators, barriers, and enablers revealed by this study, and their possible impact on the visibility of library resources in SERP. Each of the motivating, inhibiting, and enabling factors discussed in this section merits further study. The above analysis showed that these factors are interconnected. Barriers feed into and reinforce each other; enabling factors may strengthen motivations, and motivations may encourage libraries to develop enabling factors. This interconnectedness can make it difficult to know where to begin in mitigating barriers to visibility. However, addressing negative attitudes about search engines and mitigating technical difficulties were suggested as two logical places to start unraveling this complex knot of barriers.

CHAPTER 6 – CONCLUSION

6.1 AN OVERVIEW OF FINDINGS

As the above discussion shows, search engine visibility for libraries is a complex and rapidly-evolving issue. An analysis of interviews with library staff reveals that there are many interconnected factors to be considered. Librarians are motivated to increase search engine visibility for library resources by a desire to adapt to the public's changing information seeking behaviour, reach out to people who are not currently library users, and contribute to the quality of the information available to Canadians through search engines.

Despite these motivations, many factors stand in the way of visibility. Barriers exist in the form of attitudes, including: negative perceptions of search engines, fears that search engine visibility is an impossible goal, fears about the unknown consequences of becoming more visible, the perception that libraries do not need added exposure on search engines, concerns about data security and privacy, and attachment to current practices and formats. Organizational issues can also act as barriers to visibility: lack of integration into organizational goals, lack of comfort with technology throughout organizations, lack of buy-in for visibility initiatives, lack of specific positions or departments that explicitly address search engine visibility, and lack of policies surrounding visibility may cause search engine visibility to fall between the cracks. Technological difficulties range from difficulty in having resources indexed by search engines, to difficulty rising high enough in search engine results to be truly visible to searchers. Some of these technological difficulties have known solutions, such as using sitemaps to direct search engine crawlers to the appropriate library resources, while others are more difficult to resolve, such as the current lack of geolocation in book-related searches.

While these barriers may be daunting, this study also revealed a number of potential enabling factors that may help libraries approach the issue of search engine visibility. Inter-library communication and cooperation emerged as a powerful tool that could help

libraries learn from each other's experiences, share insights, and negotiate with third parties. It is my hope that this study will be an initial step towards opening the lines of communication among libraries, and fostering cooperative efforts to address this complex issue. Positive attitudes about search engines and the potential of search engine visibility were also very apparent in these interviews, which offers hope that librarians will be increasingly willing to prioritize this issue. Organizational factors such as building a culture of innovation, as well as a willingness among leaders to take risks, can increase libraries' ability to adapt. Technological enablers were also identified, including: up-to-date OPAC and ILS technology, the ability to control and experiment with library technology at the library-level, and access to high-quality and relevant data about website traffic to library websites.

Is genuine, impactful search engine visibility possible for library resources? At this point, the answer to that question is unclear. More research, and more data, are necessary to determine search engines' true potential as an outreach tool for libraries. However, as discussed in Chapter 5, it is a question worth answering, and an issue worth prioritizing despite the risk of failure. Librarians may not be comfortable with every aspect of search engines as companies or as information seeking tools. On the other hand, librarians may not be comfortable with everything that happens on the streets of their communities, but that does not stop them from opening the doors of their branches. Being accessible through search engines is rather like opening the front door of a library branch - the minimum of accessibility necessary to engage with the community. Doing so does not require total agreement between libraries' values and the values and practices of search engines, just as libraries do not require total agreement between their values and the values and practices prevalent in their local communities. Rather, libraries champion their core values by participating in their communities; libraries should participate actively in the online community for the same reason.

6.2 Recommendations for Libraries

Based on the above research, the following recommendations are suggested to help libraries capitalize on the opportunities inherent in search engine visibility by tapping into the potential enablers identified in this study, and mitigating the identified barriers.

1. Incorporate search engine visibility into organizational priorities. Build responsibility for this type of visibility into job descriptions and departmental structures, in order to create space for the consideration of questions surrounding search engine visibility.

2. Prioritize staff training in the technology involved with search engine visibility. Encourage staff to make use of the information provided by search engines to gain a better understanding of the mechanics of search engine relevance and indexing.

3. Reach out to other libraries in order to learn from each other's experiences, share insights, and collaborate on solutions to barriers. It can be difficult to assess the risks and benefits of entering this relatively new territory, and there is little published literature on the topic. Informal information sharing can help libraries mitigate risks by avoiding pitfalls experienced by others, and adopting practices other have found effective.

4. Engage with vendors on the topic of search engine visibility, in order to leverage the potential for innovation in the library-vendor relationship, and ensure that vendors are aware that this issue is important to libraries.

5. Take simple, immediate steps to improve the visibility of library resources, such as ensuring that the library's catalogue has a properly configured /robots.txt file, and an up-to-date sitemap. Search engines provide detailed guidelines about how

to optimize these files.[4] If these steps are outside of the library's control, open a dialogue with the OPAC vendor to request that the vendor either performs these tasks, or provides the library with an opportunity to do so.

6.3 Suggestions for Further Research

One of the goals of this study is to address the lack of existing research about search engine visibility for Canadian public libraries by providing a groundwork for future investigation. The results of this study suggest multiple avenues for future research. For instance, one of the limitations noted in Section 3.6 is the fact that libraries at different stages of visibility encounter different factors, which causes a lack of agreement between participants on perceived barriers, enablers, and motivators. The small scope of this study makes it difficult to make any generalized conclusions about Canadian libraries. However, this study did identify a set of potential motivators, enablers, and barriers as perceived by at least two or three Canadian librarians. Based on the data reported here, the prevalence of those factors could be tested using a more widely-reaching method such as an online survey. Similarly, the weight or importance of each factor could be tested, in order to help guide libraries' visibility efforts.

Another avenue of research is suggested by the prevalence of vendors as both barrier and enabler. Because this study did not include any participants from vendor organizations, it cannot reveal any information about vendors' motivations or priorities. Collecting more data about vendors' role in the visibility of library resources would be a valuable contribution to this area of research. Similarly, it would be interesting to collect original data concerning search engines' perspective on this issue. Further investigation into search engines' intentions for semantic search and their perceptions of libraries as potential partners would be a valuable next step to build on the information provided by this study. One of the most pressing questions left outstanding by this study is, how can libraries and search engines cooperate? The Semantic Web and linked data seem to offer

[4] For example, here: https://support.google.com/webmasters/answer/6062608?hl=en and here: https://support.google.com/webmasters/answer/183668?hl=en

an opportunity for collaboration, but search engines' and libraries' efforts to realize linked data's potential do not appear to be connecting (see Section 2.2). This study has shown that the lines of communication between libraries and search engines are not currently functioning very well. Future study illuminating why this is the case, and revealing ways to open up productive lines of communication, would be helpful.

This study approaches search engine visibility for library resources as a positive goal, but more research is necessary to support the possible impact of pursuing that goal. Gathering more data about search engine visibility's impact on library patrons, as the end-users of library resources, would be a useful avenue to pursue. Because few libraries have succeeded in gaining true visibility for their resources in search engines, there is little empirical evidence to support the impact of such visibility. A detailed comparison of a library's website traffic and circulation statistics before and after having library resources indexed by search engines, for instance, would be helpful in assessing whether making library resources more visible in SERP has a genuine impact on the use of library resources. The idea that making library resources more visible would improve online searchers' access to information, expressed by participants (see Section 4.2.3) and in the literature (e.g., Thurow, 2015), is also largely untested. This gap could be remedied with a user-experience study examining how users would interact with such information if it did appear in search results. Having more evidence to support the importance of prioritizing search engine visibility could be an important factor in motivating libraries to address this important issue.

REFERENCES

"About /robots.txt." (2007). *The web robots pages.* Retrieved from http://www.robotstxt.org/robotstxt.html

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

Balatsoukas, P., & Ruthven, I. (2012). An eye-tracking approach to the analysis of relevance judgments on the web: The case of Google search engine. *Journal of the American Society for Information Science and Technology, 63*(9), 1728–1746. doi: 10.1002/asi.22707

Barker, S. (2016, Jan 10). "In age of Google, librarians get shelved." *Wall Street Journal.* Retrieved January 30, 2016 from http://ezproxy.library.dal.ca/login?url=http://search.proquest.com.ezproxy.library.dal.ca/docview/1755357108?accountid=10406

Bazeley, P. (2013). *Qualitative data analysis: Practical strategies.* Los Angeles: Sage Publications.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American,* 284(5), 28–37.

Bishop, C. A. (2012). *Access to information as a human right.* El Paso: LFB Scholarly Publishing.

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.

BrightPlanet Corporation. (November 5, 2014). How big is the internet? *Deep Web University.* Retrieved from http://www.brightplanet.com/2014/11/big-internet/

Cahill, K. & Chalut, R. (2009). Optimal results: What libraries need to know about Google and search engine optimization. *The Reference Librarian*, *50*(3), 234-247. doi: 10.1080/02763870902961969

Calaresu, M., & Shiri, A. (2015). Understanding semantic web: A conceptual model.Library Review, 64(1), 82. Retrieved from http://ezproxy.library.dal.ca/login?url=http://search.proquest.com.ezproxy.library.dal.ca/docview/1651370488?accountid=10406

Canadian Library Association. (2015). *Canadian Library Association Statement on Intellectual Freedom and Libraries.* Retrieved from http://cla.ca/wp-content/uploads/CLA_Intellectual_Freedom_Position_Stmt_27sept2015_ltrhd.pdf

Cavoukian, A. (2011). Privacy by design: The 7 foundational principles. Retrieved from https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf

Christison, A. (2013). Discovery layers and discovery services. *Catalogue and Index, 170,* p. 2-12. Retrieved from http://shura.shu.ac.uk/7435/1/AC170_article.pdf

CIHR, NSERC & SSHRC (Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council of Canada). (2014). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Retrieved from http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf

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

Council of Canadian Academies. (2015). *Leading in the digital world: Opportunities for Canada's memory institutions*. Ottawa, ON: The Expert Panel on Memory Institutions and the Digital Revolution, Council of Canadian Academies.

Cresswell, J. W. (2007). *Qualitative inquiry & research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, California: Sage Publications, Inc.

Cresswell, J.W., & Miller, D.L. (2000). Determining validity in qualitative inquiry. *Theory into Practice, 39*(3), 124-130. doi: 10.1207/s15430421tip3903_2

DBpedia. (2015). About DBpedia. Retrieved from http://wiki.dbpedia.org/about

Dempsey, L., Malpas, C., & Lavoie, B. (2014). Collection directions: The evolution of library collections and collecting. *Libraries and the Academy, 14*(3), 393–423. doi: 10.1353/pla.2014.0013

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

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

Dou, W., Lim, K. H., Su, C., Zhou, N., & Cui, N. (2010). Brand positioning strategy using search engine marketing. *MIS Quarterly, 34*(2), 261–279.

Egger-Sider, F. & Devine, J. (2005). Google, the invisible web, and librarians. *Internet Reference Services Quarterly*, *10*(3-4), 89-101. doi: 10.1300/J136v10n03_09

Enis, M. (2015a). Ending the invisible library: Linked data surfaces library holdings where users are looking online. *Library Journal, 140*(3), 36. Retrieved from lj.libraryjournal.com/2015/02/technology/ending-the-invisible-library-linked-data

Enis, M. (2015b). OCLC works toward linked data environment - ALA midwinter 2015. *Library Journal.* Retrieved from lj.libraryjournal.com/2015/02/technology/oclc-works-toward-linked-data-environment-ala-midwinter-2015

Fichter, D., & Wisniewski, J. D. (2014). Being findable: Search engine optimization for library websites. *Online Searcher, 38*(5), 74-76.

Fortier, A., & Burkell, J. (2015). Hidden online surveillance: What librarians should know to protect their own privacy and that of their patrons. *Information Technology and Libraries (Online), 34*(3), 59-72. doi: 10.6017/ital.v34i3.5495

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. Retrieved from http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf

Google. (n.d.a). Crawling & indexing: Inside search. Retrieved February 9, 2016 from https://www.Google.ca/insidesearch/howsearchworks/crawling-indexing.html

Google. (n.d.b). How search works. Retrieved February 23, 2016 from http://www.Google.com/insidesearch/howsearchworks/thestory

Google. (n.d.c). Algorithms: Inside search. Retrieved February 29, 2016 from https://www.google.com/insidesearch/howsearchworks/algorithms.html

Google. (2010). Google search engine optimization starter guide. Retrieved February 2, 2016 from http://static.Googleusercontent.com/media/www.Google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf

Google. (2015). Terms of service: Google Analytics. Retrieved from https://www.Google.ca/analytics/terms/us.html

Google. (2016a). *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

Google. (2016b). Trends graphs and forecasts. Retrieved from https://support.Google.com/trends/answer/4355164?hl=en

Google. (2016c). Use canonical URLs. Retrieved from https://support.Google.com/webmasters/answer/139066

Google. (2016d). Safeguarding your data. Retrieved from https://support.Google.com/analytics/answer/6004245

Google. (2016e). Link schemes. Retrieved from https://support.google.com/webmasters/answer/66356

Gorman, G.E. (2006). Giving way to Google. *Online Information Review, 30*(2), 97-99. doi: 10.1108/14684520610659148

Granka, L. A. (2010). The politics of search: A decade retrospective. *Information Society*, 26(5), 364-374. doi:10.1080/01972243.2010.511560

Griffiths, J. R., & Brophy, P. (2005). Student searching behavior and the web: Use of academic resources and Google. *Library Trends, 53*(4), 539-554. Retrieved from http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=16811207&site=ehost-live

Heath, T. (n.d.). Frequently asked questions. Retrieved January 28, 2016 from http://linkeddata.org/faq

Herring, M. (2005). A gaggle of Googles: Limitations and defects of electronic access as panacea. *Internet Reference Services Quarterly, 10*(3-4), 37-44. doi: 10.1300/J136v10n03_05

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2015). *NMC Horizon Report: 2015 Library Edition.* Austin, Texas: The New Media Consortium.

Kumar, S. (2011). Effect of web searching on the OPAC: A comparison of selected university libraries. *Library Hi Tech News*, *28*(6), 14-21. doi: 10.1108/07419051111173883

Leedy, P.D., & Ormrod, J.E. (2013). *Practical research: Planning and design* (10th ed.). Toronto, Ontario: Pearson.

Levy, S. (2011). *In the plex: How Google thinks, works, and shapes our lives.* New York: Simon & Schuster.

"Librarians' role changes as information does." (2016, Jan 18). *Wall Street Journal.* Retrieved January 30, 2016 from http://ezproxy.library.dal.ca/login?url=http://search.proquest.com.ezproxy.library.dal.ca/docview/1757701858?accountid=10406

Lim, J. W. A. (2009). Local touch and global reach. *Library Management, 30*(1/2), 25-34 doi: 10.1108/01435120910927493

Norris, B. (2006). Google: Its impact on the library. *Library Hi Tech News, 23*(9), 9-11. doi: 10.1108/07419050610725012

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

Osif, B. (2006). Branding, Marketing, and Fund-raising. *Library Administration & Management, 20*(1), 39. Retrieved from http://ezproxy.library.dal.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=llf&AN=502898029&site=ehost-live

Östlund, U., Kidd, L., Wengström, Y., & Rowa-Dewar, N. (2011). Combining qualitative and quantitative research within mixed method research designs: A methodological review. *International Journal of Nursing Studies, 48*(3), 369-383. doi:10.1016/j.ijnurstu.2010.10.005

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/

Reitz, J. M. (2014a). Library management system. *Online Dictionary for Library and Information Science.* Santa Barbara, California: ABC-CLIO, LLC. Retrieved from: http://www.abc-clio.com/ODLIS/odlis_l.aspx#libms

Reitz, J. M. (2014b). OPAC. *Online Dictionary for Library and Information Science.* Santa Barbara, California: ABC-CLIO, LLC. Retrieved from: http://www.abc-clio.com/ODLIS/odlis_o.aspx#opac

Ryan, G. W., & Bernard, H. R. (2003). Techniques to identify themes. *Field Methods,15*(1), 85-109.

Scheeren, W. O. (2012). *The hidden web: A sourcebook*. Santa Barbara, California: Libraries Unlimited.

SEOBook. (2016). How search works in 2016 [Infographic]. Retrieved from: http://www.seobook.com/learn-seo/infographics/how-search-works.php

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

Smith-Yoshimura, K. (2014). “Linked data survey results 1 – Who’s doing it.” *OCLC Research.* Retrieved from http://hangingtogether.org/?p=4137

Statistic Brain Research Institute. (2015). Total number of pages indexed by Google. Retrieved from http://www.statisticbrain.com/total-number-of-pages-indexed-by-Google/

Stiller, J. & Szymanski, K. (2008). Dynamic URLS vs static URLs. Retrieved 31 July, 2015 from http://Googlewebmastercentral.blogspot.ca/2008/09/dynamic-urls-vs-static-urls.html

Thurow, S. (2015). To optimize search, optimize the searcher. *Online Searcher, 39*(4), 44-48.

Uyar, A., & Aliyu, F. M. (2015). Evaluating search features of Google Knowledge Graph and Bing Satori: entity types, list searches and query interfaces. *Online Information Review, 39*(2), 197-213. doi: 10.1108/OIR-10-2014-0257

West, J. (2015, February 2). Google’s slow fade with librarians: Maybe they’re just not that into us? *The Message.* Retrieved from https://medium.com/message/Googles-slow-fade-with-librarians-fddda838a0b7#.4e7xo4amv

“Why use microdata”. (n.d.) Retrieved from https://schema.org/docs/gs.html#microdata_why

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

Zepheira, LLC. (2014a). Libhub frequently asked and common questions. Retrieved from http://www.Libhub.org/faq/

Zepheira, LLC. (2014b). Join the movement – take the Libhub Initiative pledge. Retrieved from http://www.Libhub.org

APPENDIX A: INTERVIEW QUESTIONS

1. What is your official job title?

2. Can you tell me a little bit about what your job entails?

3. Do any of your daily tasks relate to search engine indexing? By "search engine indexing" I mean the practice employed by most search engines of using automated "web crawlers" to catalogue individual web pages in order to add those pages to their index of the Internet.

4. Do any of your daily tasks relate to search engine optimization? By "search engine optimization" I mean practices and policies specifically aimed at making a website friendly to search engine indexing, in order to improve a website's visibility on search engine results pages – in order to make a website easier to find through search engines.

5. As an information professional, do you have any opinions about search engines?
    a. Can you explain that (feeling, opinion, etc) a bit more?
    b. Do you see search engines as being different from other for-profit businesses connected with libraries? (ILS providers or e-book vendors, for example?) How are they different?

6. Have you noticed any opinions about search engines among your colleagues?
    a. Why do you think people feel that way?
    b. Do you think those opinions influence the way your organization interacts with search engines?

7. How would you characterize your organization's interactions with search engines?
    a. Are search engines (or visibility on search engines) important to your organization?

b. Are search engines a factor that your organization considers when planning initiatives, etc?

8. As far as you know, are the resources listed in your Online Public Access Catalogue (OPAC) being indexed by search engines? To what extent?

9. Does your organization have explicit policies regarding search engines?
    a. Are there policies about search engine indexing?
    b. Are there policies about search engine optimization?

10. In your organization, who is responsible for creating policies regarding search engines? Who implements those policies?

11. Are there any policies or practices regarding search engines that you would like to see implemented in your organization?

12. Have you encountered any barriers which make it harder for your organization's resources to be visible on search engines?
    a. (Ask about…) Attitudes
    b. Policies
    c. Structures
    d. Technologies
    e. Other

13. Does your organization collect data on the online traffic experienced by its Online Public Access Catalogue? What portions of the catalogue?
    a. Number of hits per page
    b. Source of hits (website or link that referred users to a page)
    c. Whether hits are human or robot

14. Would you and your organization be willing to share any of that data?

15. In your opinion, does being indexed by search engines (or not) have an impact on your organization?

16. Do you think search engines pose a threat to traditional library services?

17. Would you like to see the results of my research in June, when my project is complete? If so, may I confirm your e-mail address?

APPENDIX B: FOLLOW-UP INTERVIEW QUESTIONS

1. Around this time last year, your library system was in the process of building institutional policy and best practices surrounding search engines.
    a. How is that process going?
    b. How would you describe the outcome of your efforts so far?
    c. Have you encountered any challenges?
    d. Is your library undertaking any initiatives regarding search engines right now?

2. How would you describe the current visibility of your library's resources on search engines?

3. What do you see as next steps for your library system in terms of search engine visibility?

4. Would you be willing to share any website usage data about your library's online resources?

    a. Ideally, I would love to have a look at data about the source of visitors to both your library's main site and your library's OPAC during 2015. (Optional explanation: By "source of visitors" I mean information on how visitors found each site: for example, by navigating directly to it using a bookmark, by being referred from another site, or by searching for using a search engine.)

    b. If possible, I'd also like to see the source of referrals to your library's OPAC: which websites linked visitors to the OPAC.

5. Are there any other insights you'd like to share about your (or your library's) experiences so far with search engines?

6. Would you like to see the results of my research in May, when my project is complete? If so, may I confirm your e-mail address?

APPENDIX C: RECRUITMENT LETTER, UNKNOWN PARTICIPANT

Hello,

My name is Zoe Dickinson. I'm a Master of Library and Information Studies student in the Faculty of Management at Dalhousie University. I am conducting a research project called Public Libraries, Search Engines, and the Deep Web, and I would very much like to include the XXXX library system in my research. I am interested in the extent to which public library resources listed in Online Public Access Catalogues are discoverable through search engines, and would like to explore factors which may impact libraries' decision-making processes about this issue. I hope that this project will benefit the library and information science field by gaining a better understanding of libraries' relationships with search engines, as well as identifying best practices for interaction with search engines.

In the course of this project, I will be interviewing library staff who have expertise regarding the discoverability of library collections through search engines. In particular, I am hoping to speak to library staff who manage how their library's resources (physical books, e-books and audiobooks, for example) appear in search engine results pages. Would you be willing to either direct me to the most appropriate person in your organization, or pass this e-mail along for their consideration?

I have attached a copy of the consent form, which contains more details about the project. Please let me know if you have any questions or concerns. I can be reached through e-mail or by telephone at 250-213-2129.

Thank you for your time,
[Salutation]

APPENDIX D: RECRUITMENT LETTER, KNOWN PARTICIPANT

Dear (Name of Participant),

My name is Zoe Dickinson. I'm a Master of Library and Information Studies student in the Faculty of Management at Dalhousie University. I am conducting a research project called Public Libraries, Search Engines, and the Deep Web, and hope I can interest you in taking part. I am interested in the extent to which public library resources listed in Online Public Access Catalogues are discoverable through search engines, and would like to explore factors which may impact libraries' decision-making processes about this issue. I hope that this project will benefit the library and information science field by gaining a better understanding of libraries' relationships with search engines, as well as identifying best practices for interaction with search engines.

Would you be willing to participate in this study by answering some questions in an interview (over the telephone or Skype)? The interview should take about 50 minutes of your time. [You were suggested as a possible interview participant by XXXXXX, because your role as XXXXXXX in your organization would allow you to offer a valuable perspective on my research question.] OR [Based on your role as XXXXX in your organization, I think you will offer a valuable perspective on my research question.] If you are interested, please take a look at the attached consent form for more details about the project. If you feel that you are not the right person for this interview, I would welcome your help in finding the appropriate person to speak to within your organization.

I would very much appreciate you sharing your insights. Please let me know if you are interested, or if you have any questions about the project, or about the consent form. I can be reached through e-mail or by telephone at 250-213-2129.

Thank you for your time,
[Salutation]

APPENDIX E: RECRUITMENT LETTER, PREVIOUS PARTICIPANT

Hi XXXX,

My name is Zoe Dickinson. I'm a Master of Library and Information Studies student in the Faculty of Management at Dalhousie University. In 2014, you were kind enough to help me with my research project, *Public Libraries, Search Engines and the Deep Web*. I'm continuing my research on that topic, and I would like to include a second look at the XXXX library system in my research. I am interested in any changes which may have occurred since my last project, particularly in the extent to which your library's resources listed in Online Public Access Catalogues are discoverable through search engines. I hope that this project will benefit the library and information science field by gaining a better understanding of libraries' relationships with search engines, as well as identifying best practices for interaction with search engines.

Would you be willing to participate in this study by answering some questions in an interview (over the telephone or in person)? It would take about 50 minutes of your time. Speaking with you a second time will allow me to add a longitudinal element to my study, by assessing the possible impact of your library's initiatives regarding search engines. If you feel that you are no longer the right person for this interview, I would welcome your help in finding the appropriate person to speak with in your organization.

I have attached a copy of the consent form, which contains more details about the project. Please let me know if you have any questions or concerns. I can be reached through e-mail or by telephone at 250-213-2129.

Thank you for your time,
[Salutation]

APPENDIX F: CONSENT FORM

Project Title: Public Libraries, Search Engines, and the Deep Web

Lead researcher:
Zoe Dickinson
Dalhousie University
Email: zoe.dickinson@dal.ca
Phone: (902) 494-1901

Thesis supervisor:
Dr. Mike Smit

I invite you to take part in a research study being conducted by Zoe Dickinson, who is a Master of Library and Information Studies student in the Faculty of Management at Dalhousie University in Halifax, Nova Scotia. Taking part in the research is up to you; it is entirely your choice. Even if you do take part, you may leave the study at any time for any reason. The information below tells you about what is involved in the research, what you will be asked to do and about any benefit, risk, inconvenience or discomfort that you might experience.

Please ask as many questions as you like. If you have any questions later, please contact the lead researcher.

This research looks at the relationship between public libraries and search engines. I am interested in the extent to which library resources listed in Online Public Access Catalogues are indexed by search engines, and would like to explore any factors which may impact your organization's decision-making process about this issue. By "indexed by search engines" I mean resources that are included in search engines' catalogues of online content, and therefore retrievable using search engines.

You can take part in this study if you are staff member in a public library with expertise regarding the discoverability of library collections through search engines. There will be around 10 interview participants in this study.

To help me understand the relationship between public libraries and search engines, I will ask you to answer a series of questions over telephone or Skype about your library's interaction with search engines. This should take about 50 minutes of your time. I will record, transcribe, and analyze these interviews in order to find common themes, and possible relationships between those themes. I will be using quotations from these interviews in my reports, and by agreeing to participate you are also agreeing to be anonymously quoted in my reports. I will also be looking at quantitative data such as search engine rankings in order to find any possible relationship between libraries' policies surrounding search engines and the visibility of library resources on search engines.

Participating in this study will present minimal risk to you and your organization. Because the information collected in these interviews relates directly to your job and your organization's policies, participation might pose minimal risk to your professional standing and your organization's privacy. In order to mitigate that risk, all information gathered will be stored securely to protect confidentiality, and only reported in aggregate or after anonymization. Every practical precaution will be taken to ensure that the findings of this study are not linked in any way to you or your organization. This study has the potential to benefit the library and information science field by providing a greater understanding of libraries' relationships with search engines, as well as identifying best practices for libraries' interaction with search engines. This study also offers direct benefits to you and your organization by providing an analysis of how your organization interacts with search engines, which may help inform your organization's decision-making process in future. Participating in this study will take approximately an hour and 40 minutes of your time all together.

All identifying information you give to me in the course of my research will be kept private. I will share my findings in a Masters thesis, and may also publish my findings in

the form of journal articles, posters, and presentations. Individual responses may be quoted in my reports, however all quotations will be anonymized. This means that you will not be identified in any way in my reports. In working with your information, I have an obligation to keep all research information private. All your identifying information will be kept in a separate file, in a secure place. All electronic records will be kept secure in a password-protected, encrypted file.

It is your decision whether or not you want to take part in this research project. Even if you do take part, you can leave the study at any time for any reason. There will be no negative consequences to yourself. If at any time while being interviewed you do not feel comfortable answering a question, you are free to skip that question. If you decide to stop participating at any point during the study, you can also decide whether you want any of the information that you have contributed up to that point to be removed or if you will allow me to use that information. You can also decide for up to 4 months if you want me to remove your data. After that time, it may become impossible for me to remove it because it may already be submitted as part of my thesis.

If you agree to participate in this research project, I will ask you to provide oral consent at the beginning of the interview. I am happy to share my results with you when the project is complete in May. Please contact me via e-mail if you wish to see a copy of my final reports.

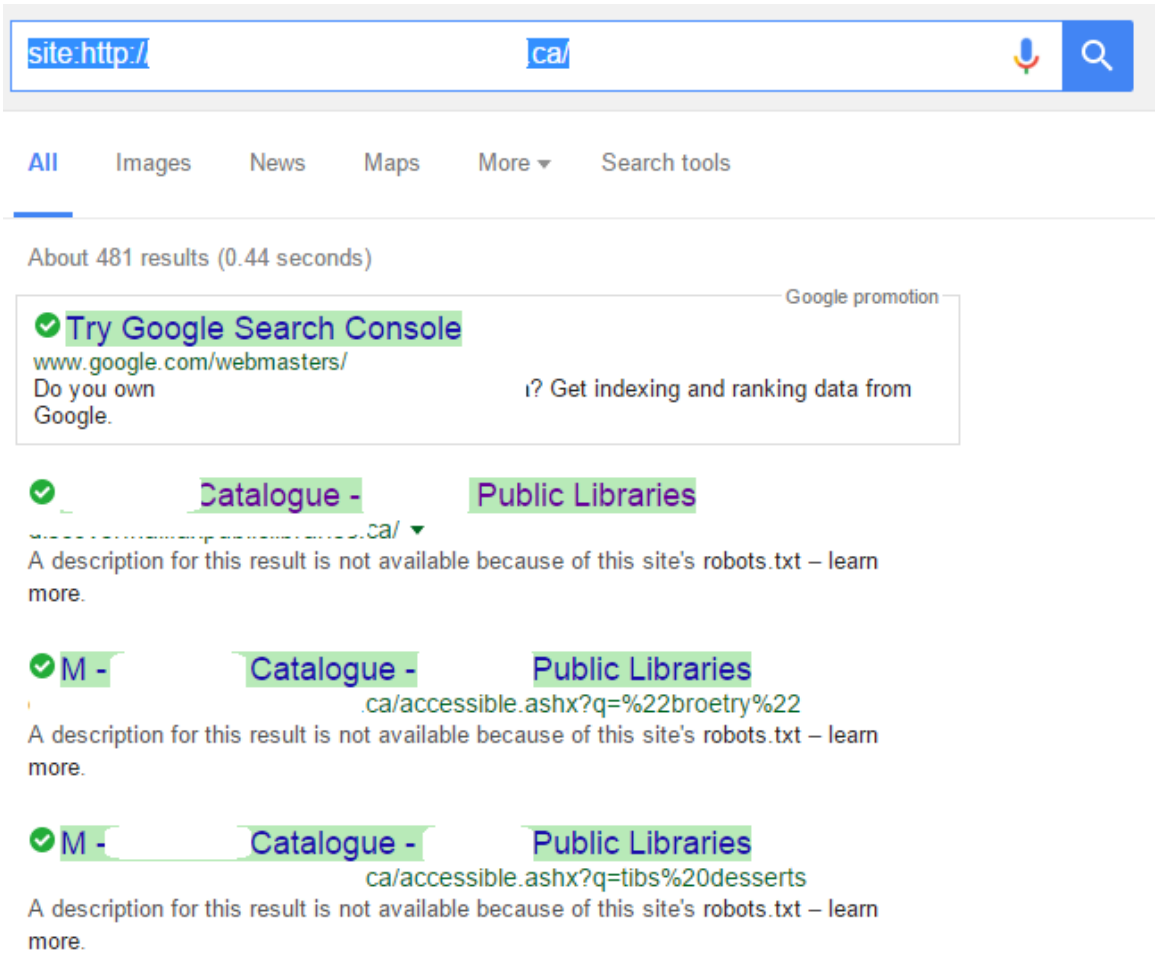
I am happy to talk with you about any questions or concerns you may have about your participation in this research study. Please contact Zoe Dickinson (at (250) 213-2129, zoe.dickinson@dal.ca) or my thesis supervisor Dr. Mike Smit (at (902) 494-1901, mike.smit@dal.ca) at any time with questions, comments, or concerns about the research study (if you are calling long distance, please call collect). I will also tell you if any new information comes up that could affect your decision to participate.

If you have any ethical concerns about your participation in this research, you may also contact the Director, Research Ethics, Dalhousie University at (902) 494-1462, or email: ethics@dal.ca

APPENDIX G: SEARCH ENGINE RESULT SAMPLES

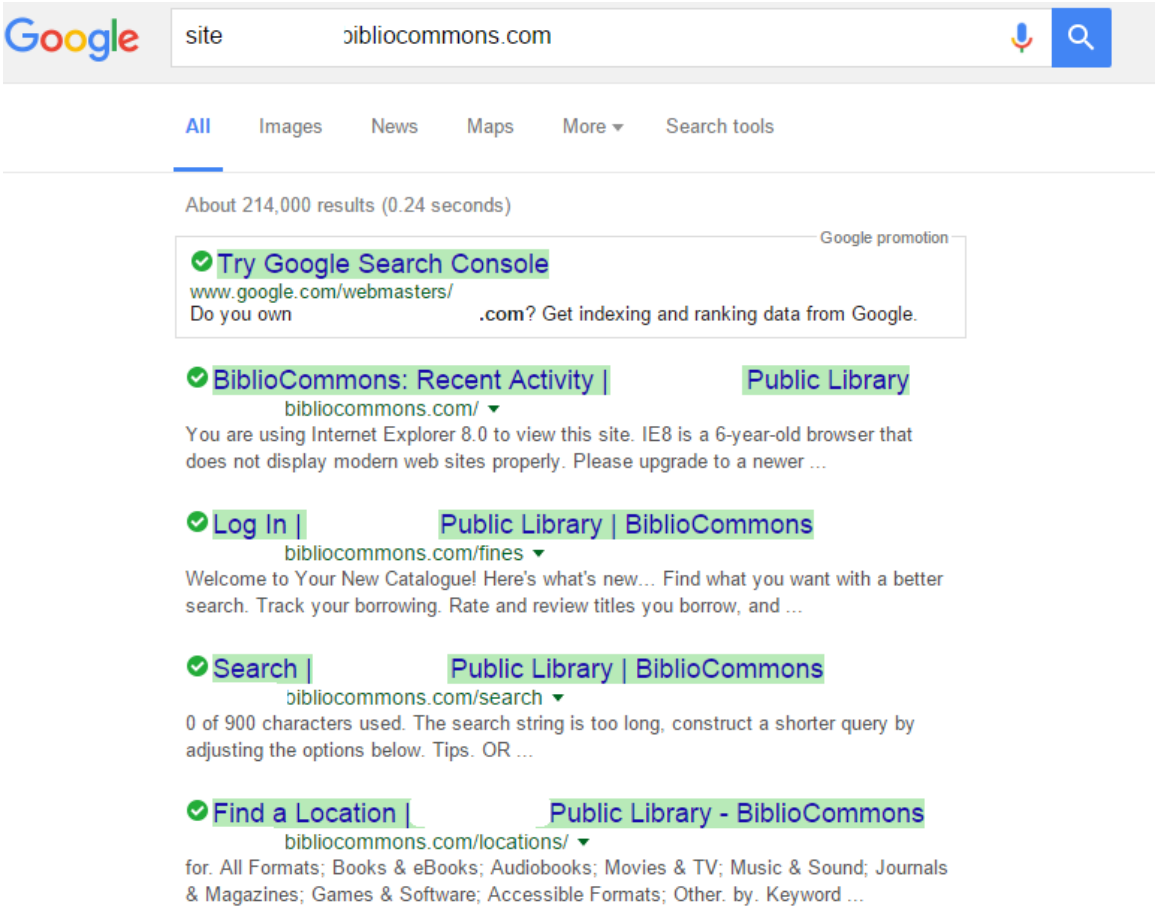
LIBRARY 1

Screenshot of SERP for L1's OPAC

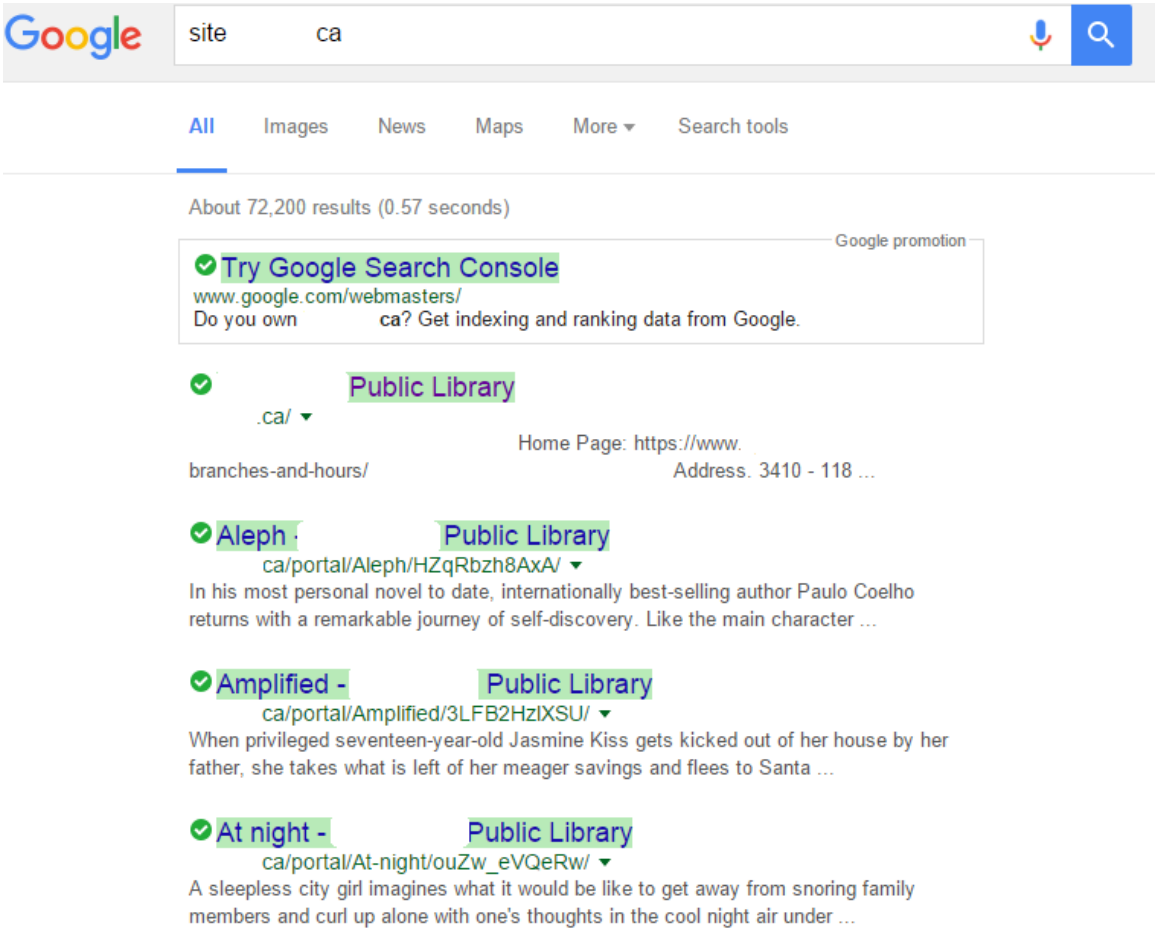


LIBRARY 2

Screenshot of SERP for L2's OPAC

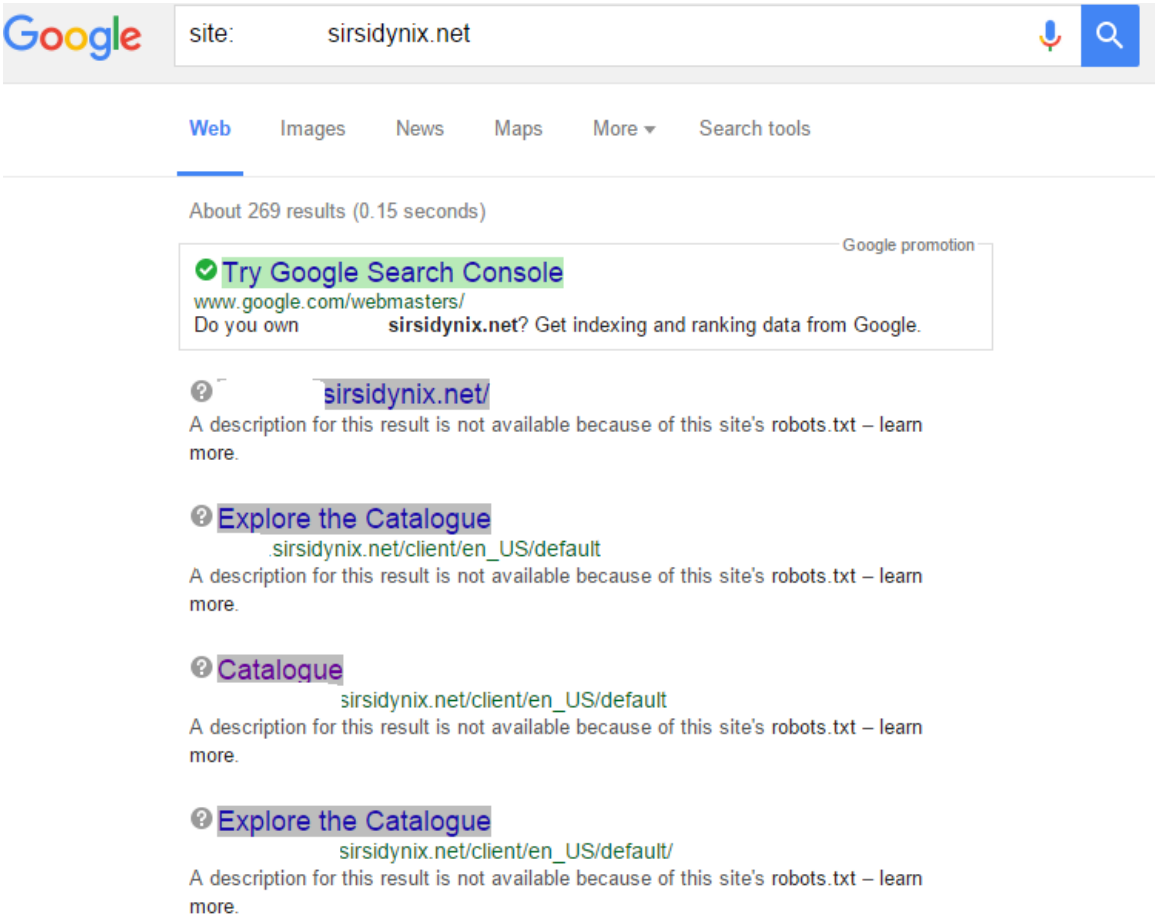


Screenshot of SERP for L2's Libhub Site



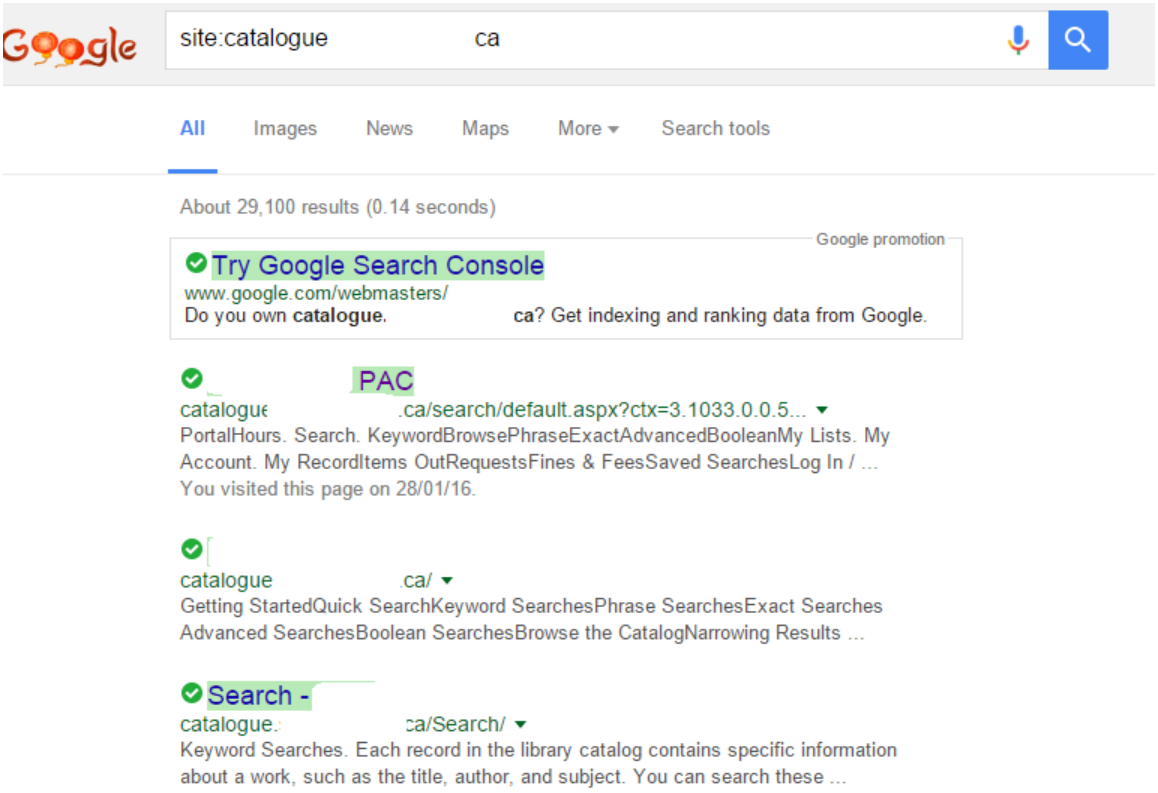
LIBRARY 3

Screenshot of SERP for L3's OPAC



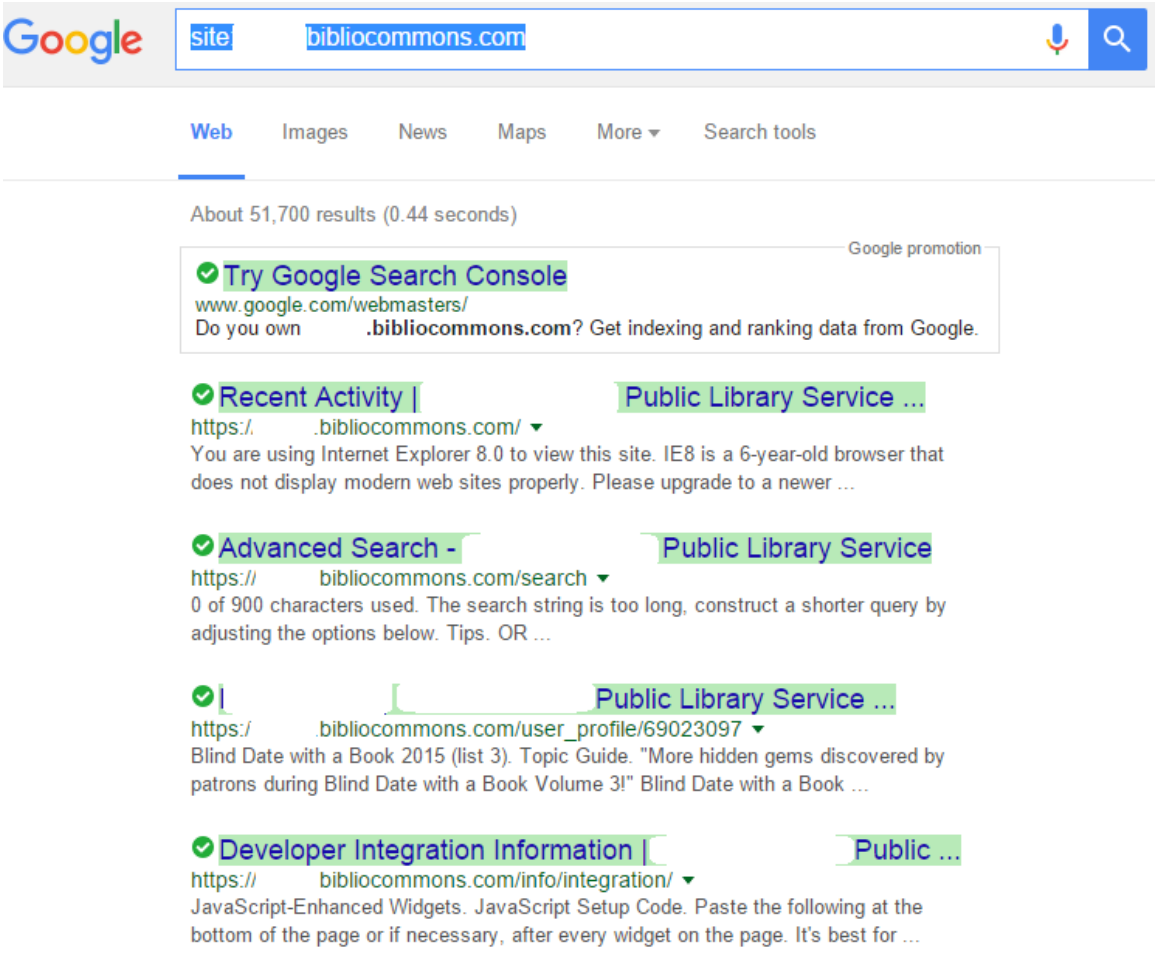
LIBRARY 4

Screenshot of SERP for L4's OPAC



LIBRARY 5

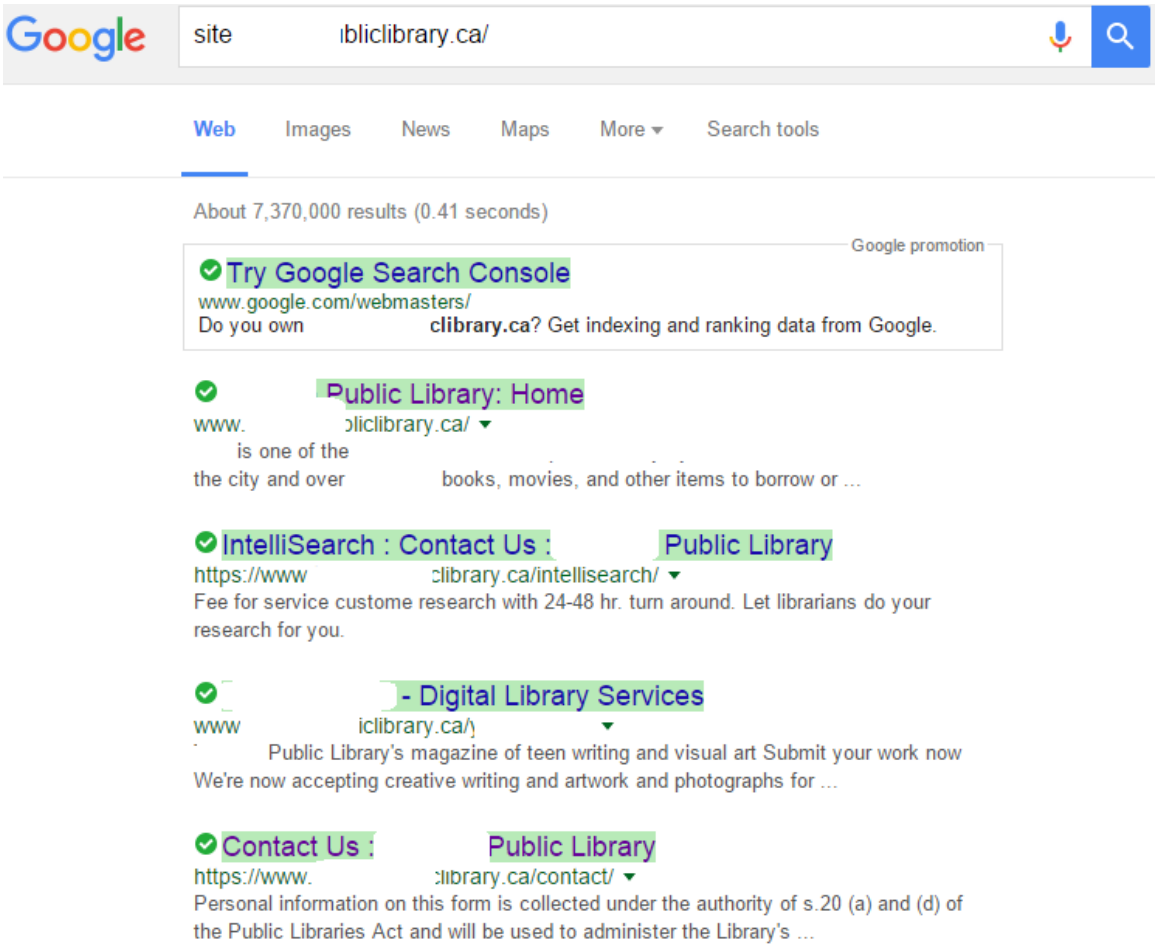
Screenshot of SERP for L5's OPAC



LIBRARY 6

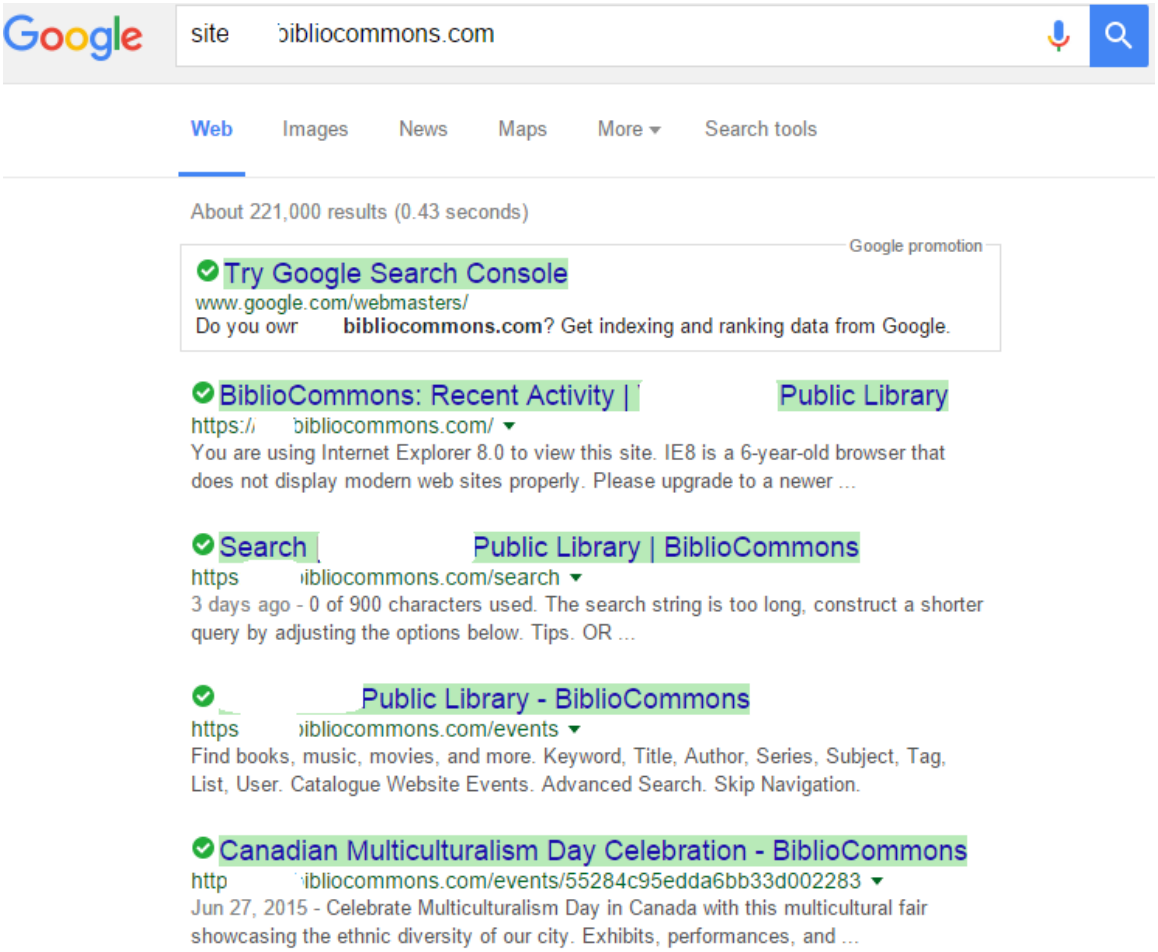
Screenshot of SERP for L6's Website

Note: Because L6 does not have a separate OPAC domain, it was not possible to perform a site-specific search of its OPAC. This result is obtained through a site-specific search of L6's main (and only) website.



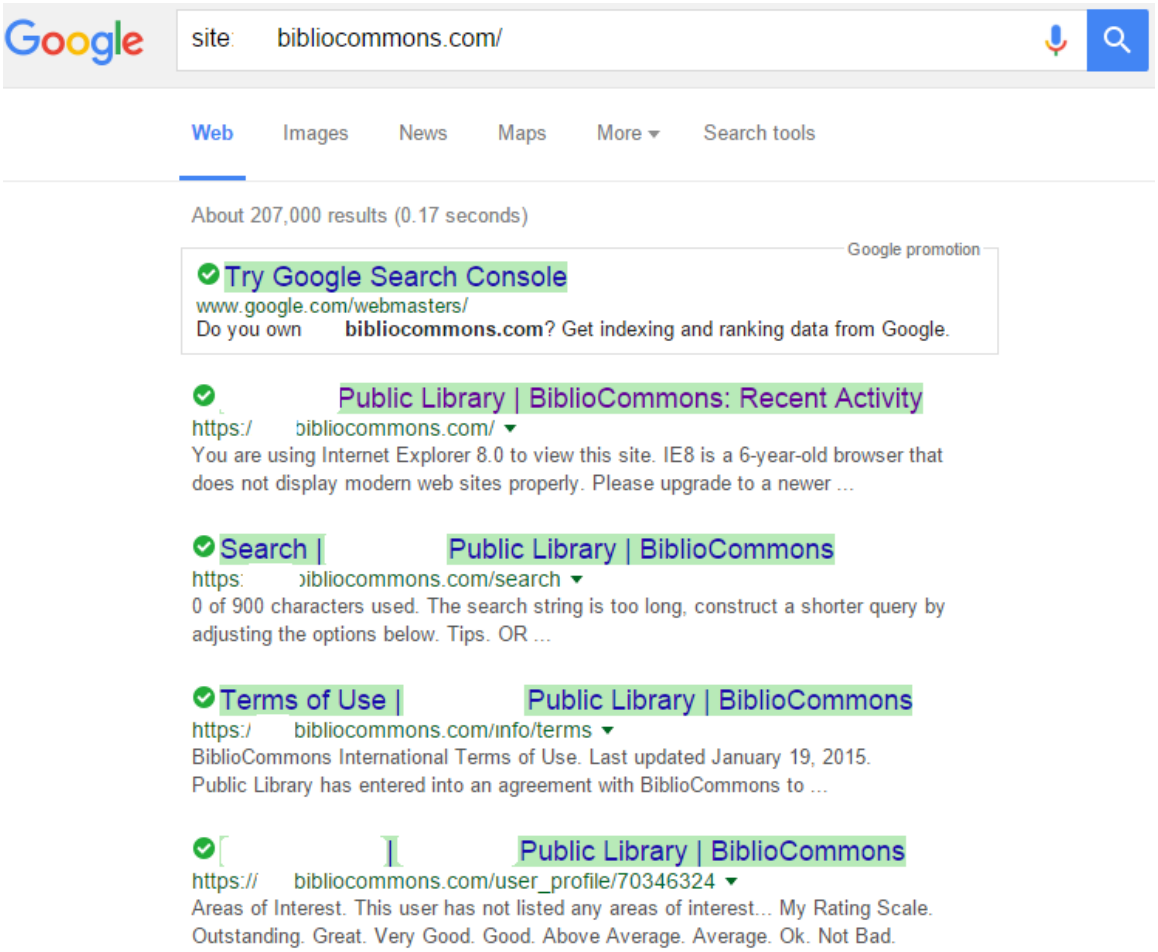
LIBRARY 7

Screenshot of SERP for L7's OPAC

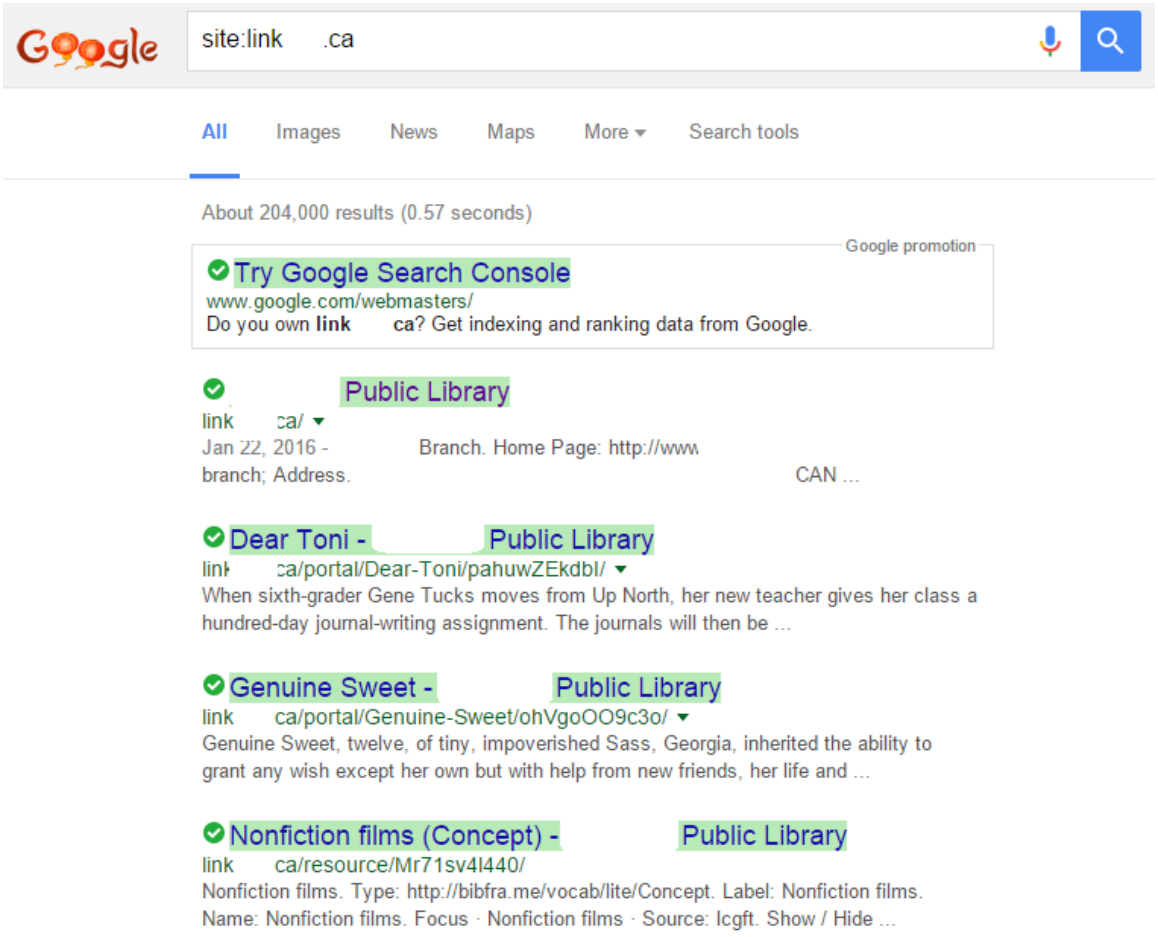


LIBRARY 8

Screenshot of SERP for L8's OPAC

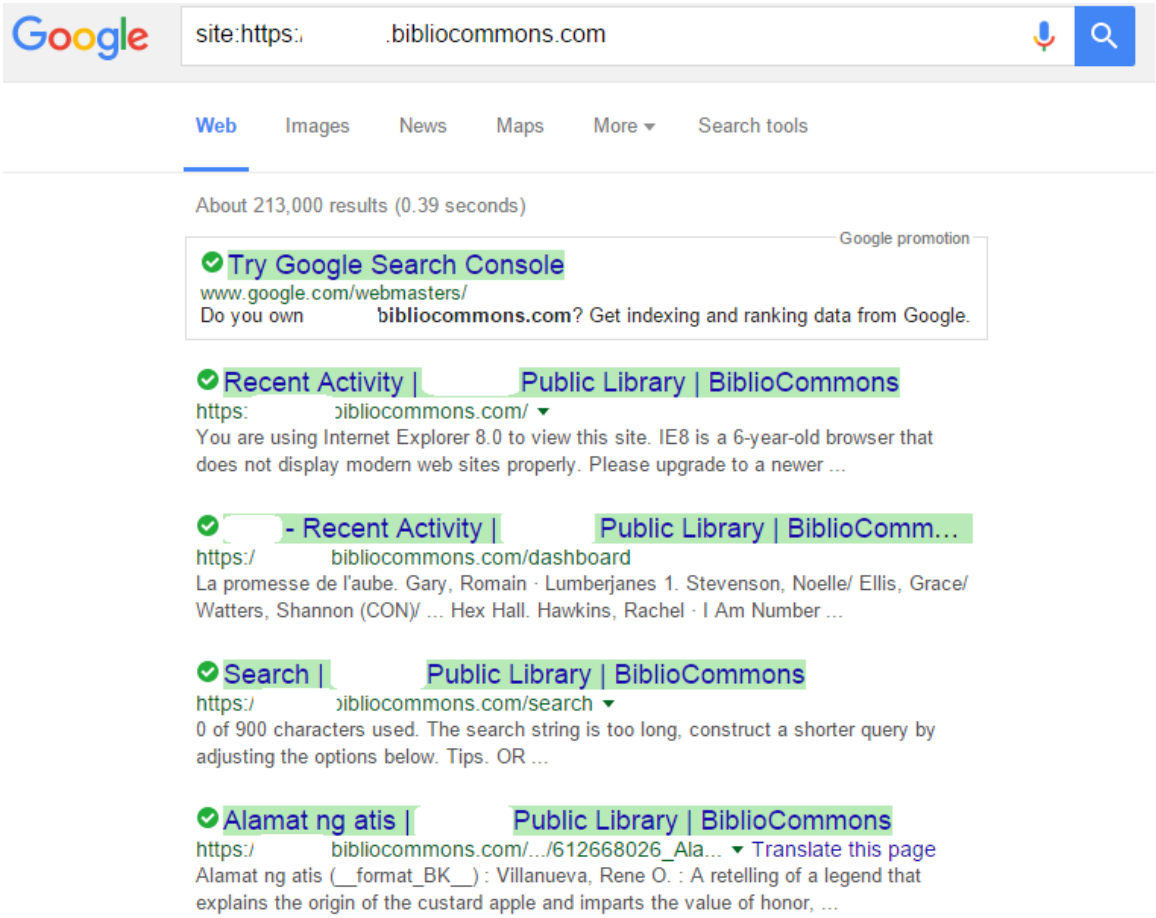


Screenshot of SERP for L8's Libhub Site



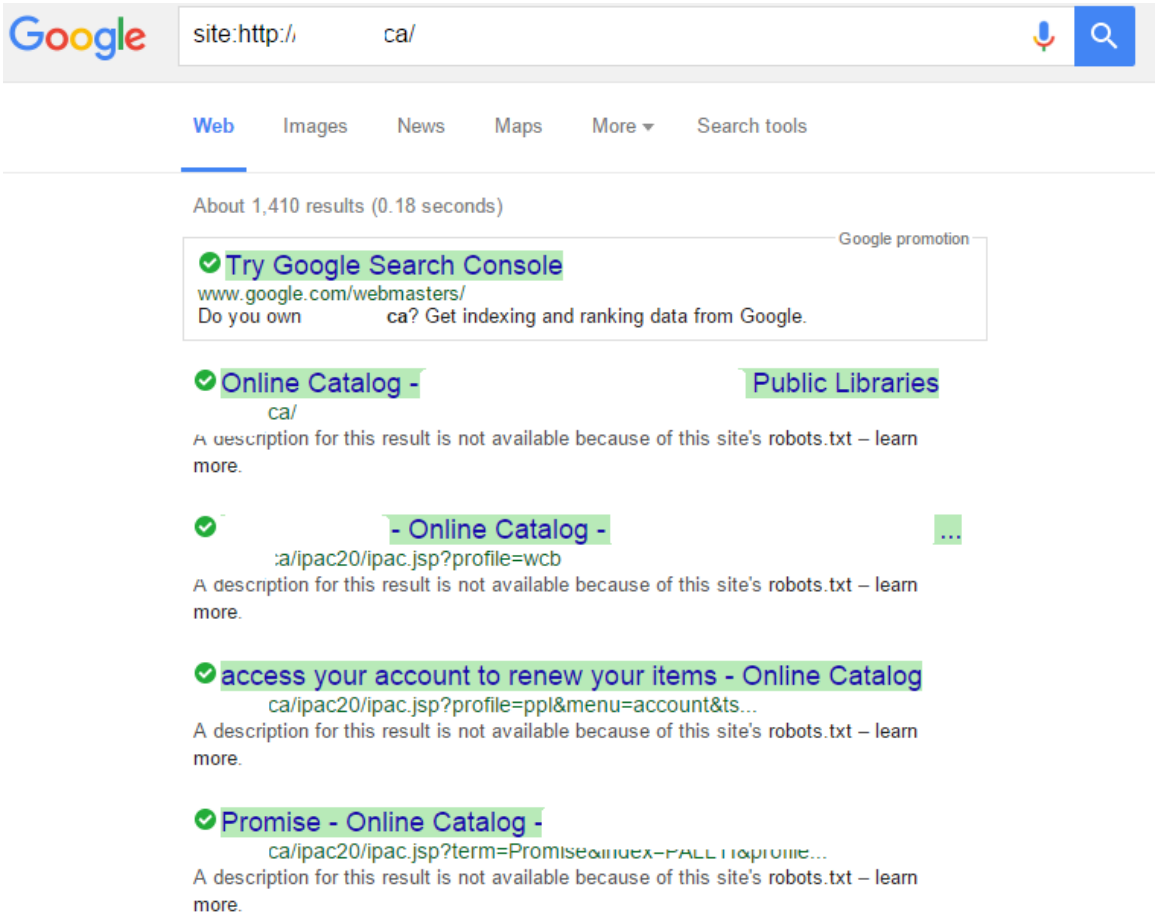
LIBRARY 9

Screenshot of SERP for L9's OPAC



LIBRARY 10

Screenshot of SERP for L10's OPAC



APPENDIX H: /ROBOTS.TXT FILES

Library 1's /robots.txt File:

```
User-agent: *
Disallow: /*
Allow: /Sitemap.ashx*
Allow: /sitemap.ashx*
Allow: /?
```

Library 2's /robots.txt File:

```
# See http://www.robotstxt.org/wc/norobots.html for documentation on how to use the robots.txt file

User-agent: *
Crawl-delay: 120
Request-rate: 1/200
Disallow: /layouts/
Disallow: /holds/select_hold/
Disallow: /item/report_match/
Disallow: /list/share/
Disallow: /info/select_library/
Disallow: /monitor/
Disallow: /item/digital_availability/
Disallow: /item/show_circulation_widget/
Disallow: /item/get_external_content/
Disallow: /item/full_record/
Disallow: /collection/add/
Disallow: /bib/match/
Disallow: /list/new/my/
Allow: /*
Visit-time: 0200-0800
```

Library 3's /robots.txt File:

```
User-agent: *
Disallow: /
```

Library 4's /robots.txt File:

```
User-agent: *
Crawl-delay: 10
# Directories
Disallow: /includes/
Disallow: /misc/
Disallow: /modules/
Disallow: /profiles/
Disallow: /scripts/
Disallow: /themes/
# Files
Disallow: /CHANGELOG.txt
Disallow: /cron.php
Disallow: /INSTALL.mysql.txt
```

```
Disallow: /INSTALL.pgsql.txt
Disallow: /INSTALL.sqlite.txt
Disallow: /install.php
Disallow: /INSTALL.txt
Disallow: /LICENSE.txt
Disallow: /MAINTAINERS.txt
Disallow: /update.php
Disallow: /UPGRADE.txt
Disallow: /xmlrpc.php
# Paths (clean URLs)
Disallow: /admin/
Disallow: /comment/reply/
Disallow: /filter/tips/
Disallow: /node/add/
Disallow: /search/
Disallow: /user/register/
Disallow: /user/password/
Disallow: /user/login/
Disallow: /user/logout/
Disallow: /eds/
# Paths (no clean URLs)
Disallow: /?q=admin/
Disallow: /?q=comment/reply/
Disallow: /?q=filter/tips/
Disallow: /?q=node/add/
Disallow: /?q=search/
Disallow: /?q=user/password/
Disallow: /?q=user/register/
Disallow: /?q=user/login/
Disallow: /?q=user/logout/
```

Library 5's /robots.txt File:

```
User-agent: *
Crawl-delay: 120
Request-rate: 1/200
Disallow: /layouts/
Disallow: /holds/select_hold/
Disallow: /item/report_match/
Disallow: /list/share/
Disallow: /info/select_library/
Disallow: /monitor/
Disallow: /item/digital_availability/
Disallow: /item/show_circulation_widget/
Disallow: /item/get_external_content/
Disallow: /item/full_record/
Disallow: /collection/add/
Disallow: /bib/match/
Disallow: /list/new/my/
Allow: /*
Visit-time: 0200-0800
```

Library 6's /robots.txt File:

```
# "Allow" and "Crawl-delay" are "non-standard"
```

```
# Not supported by all robots, but requests a 10s delay between page
loads:
User-agent: *
Crawl-delay: 10

User-agent: *
Disallow: /kidsstop/
Disallow: /eblast/
Disallow: /components/
Disallow: /config/
Disallow: /branch-computer/
Disallow: /research-computer/
Disallow: /kids-computer/
Disallow: /kids-computer-rr/
Disallow: /kiosk/
Disallow: /xml/
Disallow: /placehold
Disallow: /share-item-detail.jsp
Disallow: /rss.jsp

# WebLogic pages:
Disallow: /hou_az_ab.jsp
Disallow: /hou_az_acd.jsp
Disallow: /hou_az_ad.jsp
Disallow: /hou_az_ag.jsp
Disallow: /hou_az_ah.jsp
Disallow: /hou_az_an.jsp
Disallow: /hou_az_ap.jsp
Disallow: /hou_az_bb.jsp
Disallow: /hou_az_bc.jsp
Disallow: /hou_az_bd.jsp
Disallow: /hou_az_be.jsp
Disallow: /hou_az_bf.jsp
Disallow: /hou_az_bl.jsp
Disallow: /hou_az_br.jsp
Disallow: /hou_az_brw.jsp
Disallow: /hou_az_bur.jsp
Disallow: /hou_az_by.jsp
Disallow: /hou_az_cc.jsp
Disallow: /hou_az_ce.jsp
Disallow: /hou_az_ced.jsp
Disallow: /hou_az_ch.jsp
Disallow: /hou_az_cs.jsp
Disallow: /hou_az_da.jsp
Disallow: /hou_az_dm.jsp
Disallow: /hou_az_do.jsp
Disallow: /hou_az_dp.jsp
Disallow: /hou_az_dr.jsp
Disallow: /hou_az_dt.jsp
Disallow: /hou_az_du.jsp
Disallow: /hou_az_ea.jsp
Disallow: /hou_az_eb.jsp
Disallow: /hou_az_eg.jsp
Disallow: /hou_az_es.jsp
Disallow: /hou_az_fh.jsp
Disallow: /hou_az_fp.jsp
```

```
Disallow: /hou_az_fv.jsp
Disallow: /hou_az_ge.jsp
Disallow: /hou_az_ghp.jsp
Disallow: /hou_az_gw.jsp
Disallow: /hou_az_hb.jsp
Disallow: /hou_az_hc.jsp
Disallow: /hou_az_hil.jsp
Disallow: /hou_az_hp.jsp
Disallow: /hou_az_hs.jsp
Disallow: /hou_az_hw.jsp
Disallow: /hou_az_index.jsp
Disallow: /hou_az_jd.jsp
Disallow: /hou_az_jo.jsp
Disallow: /hou_az_js.jsp
Disallow: /hou_az_ke.jsp
Disallow: /hou_az_lb.jsp
Disallow: /hou_az_le.jsp
Disallow: /hou_az_lo.jsp
Disallow: /hou_az_ls.jsp
Disallow: /hou_az_ma.jsp
Disallow: /hou_az_mal.jsp
Disallow: /hou_az_mas.jsp
Disallow: /hou_az_mcg.jsp
Disallow: /hou_az_md.jsp
Disallow: /hou_az_mi.jsp
Disallow: /hou_az_mp.jsp
Disallow: /hou_az_mrv.jsp
Disallow: /hou_az_ms.jsp
Disallow: /hou_az_nd.jsp
Disallow: /hou_az_ne.jsp
Disallow: /hou_az_nt.jsp
Disallow: /hou_az_nycl.jsp
Disallow: /hou_az_ov.jsp
Disallow: /hou_az_pa.jsp
Disallow: /hou_az_pe.jsp
Disallow: /hou_az_pk.jsp
Disallow: /hou_az_pl.jsp
Disallow: /hou_az_pm.jsp
Disallow: /hou_az_pu.jsp
Disallow: /hou_az_pv.jsp
Disallow: /hou_az_qs.jsp
Disallow: /hou_az_rd.jsp
Disallow: /hou_az_ri.jsp
Disallow: /hou_az_rn.jsp
Disallow: /hou_az_rx.jsp
Disallow: /hou_az_sa.jsp
Disallow: /hou_az_si.jsp
Disallow: /hou_az_sj.jsp
Disallow: /hou_az_sl.jsp
Disallow: /hou_az_sp.jsp
Disallow: /hou_az_st.jsp
Disallow: /hou_az_sw.jsp
Disallow: /hou_az_sws.jsp
Disallow: /hou_az_ta.jsp
Disallow: /hou_az_th.jsp
Disallow: /hou_az_tod.jsp
Disallow: /hou_az_trl.jsp
```

```
Disallow: /hou_az_urb.jsp
Disallow: /hou_az_vv.jsp
Disallow: /hou_az_we.jsp
Disallow: /hou_az_wp.jsp
Disallow: /hou_az_ws.jsp
Disallow: /hou_az_wy.jsp
Disallow: /hou_az_yo.jsp
Disallow: /hou_az_yw.jsp
Disallow: /hou_boo_index.jsp
Disallow: /hou_loc_index.jsp
Disallow: /hou_sun_index.jsp
Disallow: /hel_cat_due.jsp
Disallow: /hel_cat_inf.jsp
Disallow: /hel_cat_man_hld.jsp
Disallow: /hel_cat_missed_holds.jsp
Disallow: /hel_cat_mke_hld.jsp
Disallow: /hel_cat_ren.jsp
Disallow: /hel_cat_sch.jsp
Disallow: /hel_cat_signin.jsp
Disallow: /hel_index.jsp
Disallow: /hel_new_catalogue.jsp
Disallow: /hel_tro_index.jsp
Disallow: /mul_index.jsp
Disallow: /mul_ser_citizenship.jsp
Disallow: /mul_ser_citizenship_fr_test.jsp
Disallow: /mul_ser_citizenship_test.jsp
Disallow: /pro_index.jsp
Disallow: /pro_nycl.jsp
Disallow: /pro_trl.jsp
Disallow: /rec_clu_index.jsp
Disallow: /rec_index.jsp
Disallow: /spe_lea_computer.jsp
Disallow: /spe_ser_index.jsp
Disallow: /spe_ser_photo.jsp
Disallow: /spe_ser_rent.jsp
Disallow: /spe_ser_wir.jsp
Disallow: /sup_index.jsp
Disallow: /sup_vol_index.jsp


# Wikipedia bot exclusions that seem like a good idea:
User-agent: Mediapartners-Google*
Disallow: /

User-agent: UbiCrawler
Disallow: /

User-agent: DOC
Disallow: /

User-agent: Zao
Disallow: /

User-agent: sitecheck.internetseer.com
Disallow: /

User-agent: Zealbot
```

```
Disallow: /

User-agent: MSIECrawler
Disallow: /

User-agent: SiteSnagger
Disallow: /

User-agent: WebStripper
Disallow: /

User-agent: WebCopier
Disallow: /

User-agent: Fetch
Disallow: /

User-agent: Offline Explorer
Disallow: /

User-agent: Teleport
Disallow: /

User-agent: TeleportPro
Disallow: /

User-agent: WebZIP
Disallow: /

User-agent: linko
Disallow: /

User-agent: HTTrack
Disallow: /

User-agent: Microsoft.URL.Control
Disallow: /

User-agent: Xenu
Disallow: /

User-agent: larbin
Disallow: /

User-agent: libwww
Disallow: /

User-agent: ZyBORG
Disallow: /

User-agent: Download Ninja
Disallow: /

User-agent: wget
Disallow: /

User-agent: grub-client
```

```
Disallow: /

User-agent: NPBot
Disallow: /

User-agent: WebReaper
Disallow: /
```

Library 7's /robots.txt File:

```
User-agent: *
Crawl-delay: 120
Request-rate: 1/200
Disallow: /layouts/
Disallow: /holds/select_hold/
Disallow: /item/report_match/
Disallow: /list/share/
Disallow: /info/select_library/
Disallow: /monitor/
Disallow: /item/digital_availability/
Disallow: /item/show_circulation_widget/
Disallow: /item/get_external_content/
Disallow: /item/full_record/
Disallow: /collection/add/
Disallow: /bib/match/
Disallow: /list/new/my/
Allow: /*
Visit-time: 0200-0800
```

Library 8's /robots.txt Files:

OPAC:

```
User-agent: *
Crawl-delay: 120
Request-rate: 1/200
Disallow: /layouts/
Disallow: /holds/select_hold/
Disallow: /item/report_match/
Disallow: /list/share/
Disallow: /info/select_library/
Disallow: /monitor/
Disallow: /item/digital_availability/
Disallow: /item/show_circulation_widget/
Disallow: /item/get_external_content/
Disallow: /item/full_record/
Disallow: /collection/add/
Disallow: /bib/match/
Disallow: /list/new/my/
Allow: /*
Visit-time: 0200-0800
```

Libhub:

```
/robots.txt:
```

```
User-agent: *
Disallow:
Sitemap: http://x.x.ca/harvest/sitemap.xml
```

Library 9's /robots.txt File:

```
User-agent: *
Crawl-delay: 120
Request-rate: 1/200
Disallow: /layouts/
Disallow: /holds/select_hold/
Disallow: /item/report_match/
Disallow: /list/share/
Disallow: /info/select_library/
Disallow: /monitor/
Disallow: /item/digital_availability/
Disallow: /item/show_circulation_widget/
Disallow: /item/get_external_content/
Disallow: /item/full_record/
Disallow: /collection/add/
Disallow: /bib/match/
Disallow: /list/new/my/
Allow: /*
Visit-time: 0200-0800
```

Library 10's /robots.txt:

```
#####     This file will tell search engines what they should index.
#####
############# Settings for all crawlers/spiders.
User-agent:*
#Don't index ANYTHING!
Disallow: /
```

APPENDIX I: TIMING OF SITE-SPECIFIC SEARCHES

Domain	Date of Search
L1 OPAC, first search	February, 2015
L1 OPAC, second search	December, 2015
L2 OPAC, first search	February, 2015
L2 OPAC, second search	December, 2015
L2 Libhub site	December, 2015
L3 OPAC	July, 2015
L4 OPAC	February, 2016 [5]
L5 OPAC	September, 2015
L6 site (there is no separate OPAC domain)	October, 2015
L7 OPAC	October, 2015
L8 OPAC	October, 2015
L8 Libhub	February, 2016 [6]
L9 OPAC	November, 2015
L10 OPAC	November, 2015

[5] In the case of L4, the site-specific search was performed later to coincide with the quantitative data provided by the participant. P4 suggested waiting until 2016 to collect quantitative data because at the time of the interview L4 was in the midst of technological transition that might impact website traffic and search engine indexing.
[6] L8's Libhub site was not originally included in site-specific searches, but was added later as a relevant data point.