Canadian public libraries and search engines: barriers to visibility

Zoe Dickinson and Michael Smit
Dalhousie University, Halifax, Canada

Received 16 September 2015
Revised 31 May 2016
Accepted 23 June 2016

Abstract

Purpose – The purpose of this paper is to examine the challenges presented by search engine visibility for public libraries. The paper outlines the results of a pilot study investigating search engine visibility in two Canadian public libraries.

Design/methodology/approach – The study consists of semi-structured interviews with librarians from two multi-branch Canadian public library systems, combined with quantitative data provided by each library, as well as data obtained through site-specific searches in Google and Bing. Possible barriers to visibility are identified through thematic analysis of the interviews.

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

Research limitations/implications – This paper describes a small, preliminary pilot study. More research is needed before any firm conclusions can be reached.

Practical implications – A review of the literature shows the increasing importance of search engine visibility for public libraries. By delving into the underlying issues which may be affecting libraries' progress on the issue, this paper may help inform libraries' decision-making processes and practices.

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

Keywords Optimization, Libraries, Visibility, Discoverability, Google, Search engine

Paper type Research paper

Introduction

Libraries and search engines

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

According to a 2010 Online Computer Library Center (OCLC) survey, 84 per cent of online information seekers began their search using a search engine; 0 per cent began with a library website (DeRosa *et al.*, 2011). In a 2012 PEW internet study, 91 per cent of adult search engine users reported that they "always" or "almost always" found the information they were looking for using search engines; only 9 per cent combined reported "only sometimes" or "hardly ever" (Purcell *et al.*, 2012, p. 14). This shows not only widespread use, but widespread confidence in search engines as an information source. As Halavais (2013) puts it, "The modern search engine has taken on the mantle of what the ancients of many cultures thought of as an oracle: a source of knowledge about our world and who we are"; or, more bluntly, "they have become an object of faith" (p. 1).

Aslib Journal of Information Management
Vol. 68 No. 5, 2016
pp. 589-606

2050-3806
DOI 10.1108/AJIM-09-2015-0147

The deep web

Despite these positive public perceptions, search engines barely begin to encompass all the information available online. As of 2009, 95 per cent of the web, representing over 220 billion pages, was not indexed by search engines (Scheeren, 2012). This is the deep web. The resources listed in libraries' online public access catalogues (OPACs) used to inhabit the deep web due to technical limitations, but as databases evolve and search engine crawlers improve, this hidden state is no longer a technological necessity. Nevertheless, the contents of most Canadian public libraries' OPACs remain hidden in the deep web (Blandford, 2015). While most libraries have an online presence, few are visible online unless a user is specifically searching for a library. If a user simply searches for an information resource, the library remains invisible even if it offers the resource in question.

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. By failing to integrate their information resources with search engines, libraries risk being invisible to the people who need their services. This endangers not only the perceived relevance of libraries (Blandford, 2015), but also the average person's ability to access quality information. If public libraries, as champions for public information access, allow themselves to drop out of sight in today's most popular information forum, people may not be aware of the essential services they offer.

There has been discussion of the relationship between libraries and search engines in library and information science (LIS) literature, but less has been done to address the question of why most libraries' resources remain invisible to search engine users. If search engine visibility is technically possible, why it is not happening? Throughout this paper, the term "search engine visibility" is used specifically in relation to the visibility of the resources listed in libraries' OPACs; the visibility of libraries' main websites and other online activities is outside the scope of this study. "Visibility" is used to describe both the extent to which an OPAC's contents are being indexed by search engines (and thus, have the possibility of being visible in search engine results pages (SERP)), and the extent to which an OPAC's contents are rising high enough in SERP to be realistically visible to an average searcher.

This paper outlines the findings of an exploratory pilot study investigating the search engine visibility of two Canadian public library systems. Potential barriers to visibility were identified through qualitative analysis of interviews with a library staff member from each institution, while the actual visibility of each institution's resources was assessed through analysis of website traffic statistics and site-specific searches on Google and Bing. These findings are detailed further in the following sections: literature review, methodology, limitations, results, discussion, and conclusion.

Literature review

Search engines as the competition

One possible explanation for libraries' continued invisibility on search engines can be found in a review of LIS literature. As Blandford (2015) notes, many LIS scholars approach search engines as a threat. Some argue that the limitations of search engines make librarians more necessary than ever as champions of information literacy (e.g. Norris, 2006; Herring, 2005; Cahill and Chalut, 2009; Egger-Sider and Devine, 2005). Although these authors emphasize search engines' limited results, and discuss the deep web, there is no mention of the fact that this group of hidden resources often includes the library's own holdings. Others suggest that libraries should compete by emulating

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

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 on SERP.

Advocates for SEO argue that libraries must integrate their resources with the larger "information ecosystem" (Arlitsch, 2014, p. 610), and see search engines as a "funnel for channeling patrons back to the library" (Breeding, 2014, p. 26). Thurow (2015) 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). 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 (Breeding, 2014). In addition, change is necessary in both attitudes and departmental structures; rather than relegating SEO to the IT department, administrators must integrate SEO with their organization's overall mission in order to ensure staff at all levels are aware of its importance in reaching the community (Arlitsch's *et al.*, 2013).

While many commercial business owners would consider the benefits of SEO obvious, in the library community 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 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 it receives" (Onaifo and Rasmussen, 2013, p. 102). Tony Boston (2005) 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. Between 2002, when the pictures catalogue was first exposed to search engine robots, and 2004, the collection saw an increase in use of 370 per cent (Boston, 2005). Marshal Breeding (2006), one of the most prolific proponents of SEO, 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. More recently, Lee *et al.* (2016) implemented a set of search engine recommended SEO techniques to LG Sangnam Library's Science Land digital repository and reported a significant increase in the percentage of visitors reaching their site via search engines: from 39.19 per cent of the site's visitors in March, 2014 to 61.57 per cent in May, 2015.

These studies hint at the possibilities offered by SEO, but a true assessment of the impact search engine visibility might have is hindered by the widespread lack of that visibility throughout the library community, particularly when it comes to physical rather than digital collections. Until more library content begins showing up in SERP, it is difficult to predict what impact that visibility might have.

Linked data

More recently, attempts are being made to connect library content with the rest of the internet using linked data, also known as the semantic web. Linked data attempts to help computers understand the context and semantic meaning of data, allowing computers to automatically link data together in more meaningful, productive ways. For example, major search engines such as Google and Bing have begun providing more relevant and exact answers to search engine queries, using ever-expanding linked data databases (Uyar and Aliyu, 2015). They also encourage linked data use by content providers through schema.org, a collaborative initiative to build structured data standards online. However, schema.org's all-purpose markup lacks granularity, overlooking concepts that are important to libraries (Enis, 2015b). As a result, several library-centric linked data initiatives have sprung up.

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. Currently, most libraries use machine readable cataloguing (MARC) standards to structure the data connected with each resource in their catalogues. This standard, 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 new 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, 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, 2014b).

While the Library of Congress has been working on BIBFRAME as an alternative to both MARC and schema.org's less detailed data structure, another group under the name of Schema Bib Extend has been working instead to expand and improve schema.org for library use. This group, which includes the 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 and 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 and Denenberg, 2015, p. 4). The hope of both initiatives is that they will eventually complement each other, with BIBFRAME serving as a new cataloguing standard for libraries, and schema.org serving as a more generally usable "data aggregator to import and export BIBFRAME data" (Godby and Denenberg, 2015, p. 8).

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 timeline 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.

Gaps in the literature

The literature to date 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 (2013) showed a majority of Ontario libraries with quite low numbers of indexed pages. What stands between these resources and search engines? Arlitsch comes closest to offering an answer to this question, in noting libraries' attachment to forms of metadata which are not compatible with search engines (Arlitsch, 2014), and identifying a lack of SEO awareness among library staff (Arlitsch's *et al.*, 2013). However, these are personal observations, rather than empirical evidence. The above literature review also suggests one possible explanation: the fact that many librarians approach search engines as a threat, rather than as useful outreach tools. If Arlitsch, 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 obstacles are stopping libraries from reaching out to patrons using today's most popular information retrieval forum? The first step in dispelling these barriers is identifying them.

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. Online visibility, as a global form of outreach, is seen as necessary for digital resources, but less relevant to physical resources such as books or library programmes. Indeed, the majority of the literature cited above relates to digital, rather than physical, collections. This dichotomy has led libraries' SEO initiatives to be focused on digital materials, rather than physical collections.

A digital-collections 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. Standing on an unfamiliar street corner, one can Google "pizza" on a smartphone and find the nearest local vendors. Similarly, one should be able to search for a book and find the nearest copy, which may very well be 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.

Methodology

In order to address these gaps in the literature, a pilot study was performed to investigate the following research question:

RQ1. What barriers stand between Canadian public libraries and search engine visibility?

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 towards search engines and SEO; quantitative studies could be designed to test searchers' potential reactions to seeing more library resources in SERP. However, because the research question involved many unknowns, and there were so few pre-existing concrete data points, qualitative interviews were chosen as the first step in illuminating this issue. Without 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.

As an initial pilot study, this research focused on two major Canadian public library systems. The two libraries chosen were both large, multi-branch systems based in major Canadian cities. At the time of the study, Library 1 had 14 branches, while Library 2 had 19. The libraries were chosen based on availability of eligible interview participants, as well as comparability in size and type between the two libraries. The study was limited to Canadian libraries in order to ensure that data would be comparable between the two libraries, with the hope that further research could eventually include a broader, more international scope.

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 the authors to explore themes as they arose in interviews, uncovering potential barriers to search engine visibility. Quantitative data were used to triangulate and strengthen the validity of the themes that arose during qualitative analysis. 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 *et al.*, 2011).

Qualitative data were assessed in the form of semi-structured interviews with library staff in charge of each library's online presence. An interview participant was chosen for each library in consultation with the organization in question, based on candidates' current roles within each organization. Interviews lasted approximately one hour each, and were digitally recorded before being transcribed and anonymized.

Each interview was analysed using thematic analysis techniques, in order to uncover themes, and discover the relationships between themes (Bazeley, 2013; Ryan and Bernard, 2003). A tentative set of broad initial themes was gleaned from the above literature review. Based on various LIS authors' speculations about why library resources remain less visible in SERP, the interviews were initially examined for potential barriers to visibility in the form of policies, organizational structures, and technical difficulties. The current discourse on the topic also prompted a focus on themes involving librarians' perceptions of search engines, in order to confirm or deny speculation about negative attitudes surrounding search engines as a potential barrier. Through close readings of each transcript, additional codes were developed to describe concepts that recurred throughout the two interviews. Upon completion of the study, participants were given an opportunity to comment on the results, suggest clarifications, and confirm content validity.

The current visibility of each library was assessed using quantitative data. The number of OPAC resources currently being indexed by search engines was assessed by performing site-specific searches in Google and Bing. These searches, while not exact, gave an approximate count of the number of pages being indexed for each OPAC. This method also allowed for an analysis of the quality of the search results being generated by each library, by producing search results as they would appear to search engine users. Examining the quantity and quality of results revealed the actual visibility of each library. In addition, website usage data were analysed in order to investigate how users were reaching each library's OPAC. Because the two libraries used different vendors for their OPACs, slightly different types of data were available in each case. This made it impossible to directly compare each library's usage statistics. However, the data provided by each library did allow for individual assessments of usage patterns surrounding search engines.

Limitations

This study was undertaken as an exploratory pilot project. Rather than attempting to provide broadly applicable data, the authors' intent was to establish an effective methodology to be expanded upon in further research, and to spark scholarly conversation on this topic. Thus, this study's scope is quite narrow, focusing on Canadian public libraries only, and considering only two of many potential Canadian public libraries in its analysis. The authors of this paper are currently conducting a larger study including ten Canadian public libraries, using the methodology tested here as a foundation.

The major limitation of this study is its small sample size. Speaking to only two participants limits the study's external validity; it cannot be assumed that the two participants provide a representative sample of all Canadian librarians, or that the two libraries in question are representative of all Canadian library systems. Thus, it is not possible to draw conclusions about the wider world of Canadian libraries from the data collected here, let alone public libraries in general. In fact, it is possible that the participants are not even representative of librarians within their own institution. Another member of the same institution might give different answers.

Despite these limitations, as an initial exploratory study, this project provides useful information to help guide more in-depth research on the topic. In addition, while the results of this study cannot be extrapolated to apply more broadly, they do show that the observed barriers exist for at least one staff member of at least one library system in Canada. This information in itself is valuable, but further study will be necessary before any broader conclusions can be drawn.

Another limitation of this study is in its inability to draw direct links between qualitative and quantitative data. Although correlations can be suggested between participants' perceptions and the quantitative data collected by their libraries, no true causal links can be made. There are many other possible factors which may be affecting the number of pages indexed from each OPAC, and the website usage statistics collected by each library. Taken together, the qualitative and quantitative data help illuminate the state of each library's relationship with search engines. The study design included the collection of a predetermined set of quantitative data to minimize the "ask" of participants. The inability to collect additional quantitative data based on themes uncovered in the interviews limits the connections that can be made between the qualitative and quantitative data. These connections should be explored in future studies, as there is not currently enough evidence to establish any causal links.

Results: interviews

Roles

Both participants were responsible for the online presence of their respective libraries, but the similarity in roles ends there. As web developer, Participant 1's responsibilities began and ended with Library 1's main website. Although he dealt with the OPAC periodically in order to populate the main website with content, the web presence of the OPAC itself did not fall within his domain. Because his duties did not encompass the OPAC's online visibility, Participant 1 was not in an ideal position to answer questions related to the visibility of library resources on search engines. However, from his description of Library 1's organizational structure, it appears that no position existed which dealt directly with that issue. To his knowledge, "there is no SEO on our ILS [integrated library system, expressed online through an OPAC]", and the team responsible for the ILS and its front-facing OPAC did not include SEO in their duties. Participant 1 seemed to be the closest possible match for this study's requirements available within Library 1.

In contrast, Participant 2's role was "mainly about that word "discovery" which is very hot in libraries right now, and it's about the idea of helping people naturally discover online resources, either to use them more fully or to discover new things about them". Her focus was on the "digital display of collections and services", encompassing both the main website and the OPAC. As a part of the collections department, Participant 2 was working closely with members of the web redevelopment team, ILS admins, and cataloguing librarians to improve Library 2's online visibility. At the time of the interview, she was investigating how many of the resources in its OPAC were being indexed by search engines, and how visible those resources were to the average searcher.

Attitudes

The two participants expressed very different attitudes about search engines. Participant 1 raised concerns about Google's motivations which, combined, presented a barrier to search engine indexing of his library's resources. Participant 2, on the other hand, had overall positive feelings about search engines.

Participant 1 expressed serious misgivings about search engines in general, and Google in particular as the dominant player in the search industry:

> Google is […] a business that cares more about making money and holding the information monopoly of the world, rather than helping people find books in a library […].

In addition, while Participant 1 did not see search engines as an immediate threat to his specific library or its catalogue, he did identify "an overarching threat to libraries, because they're taking over some of the traditional roles that librarians have done". These negative feelings made Participant 1 wary of allowing a search engine to index his library's OPAC:

> 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.

From this perspective, visibility of resources on search engines does not align with the mission of libraries, and therefore is not seen as a desirable goal.

Participant 2, on the other hand, viewed search engines as an integral part of libraries' current goals: "a very big goal of libraries is to finally have their data talk to search engines". She saw this goal as being in alignment with the fundamental purpose of public libraries: "As a public librarian, […] you want to be where people are and you want to serve them as best as possible". According to Participant 2, being visible on search engines is important to that purpose because "at this point, I think it's pretty widely acknowledged that search engines are the place that people are performing their information seeking behaviour".

Rather than seeing search engines as a threat to libraries, she saw them as "very powerful tools that are incredibly relevant to our customers". To her, these tools were an important way for libraries to fulfil their purpose: "And so I think search engines are definitely a way to help serve the people as best as possible".

While recognizing the benefits that search engines offer to libraries, she also pointed out the benefits that libraries can offer search engines:

> 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.

In this mutually beneficial arrangement, search engines would help libraries reach patrons, and libraries would help search engines improve the quality of information they offer.

The two participants' responses were equally disparate regarding opinions about search engines among their colleagues. Participant 1 described a general lack of opinions about, or interest in, search engines in his organization. This lack of interest in itself may be a barrier to visibility on search engines, as the issue may not be considered or discussed. Participant 2, on the other hand, described a great deal of interest in "making ourselves as visible as possible". This interest is evident in the fact that Participant 2's organization is an early adopter of the Libhub initiative (see the "Linked Data" section of the Literature Review), with a three-department task force working on shifting the way the library deals with its resource records, in order to "help search engines find our data in the most effective way possible".

Policies

Organizational policies concerning search engines were a point of commonality between the participants. Both participants' libraries did not have any written policies about search engine indexing or SEO. In both cases, the participants themselves were in the process of performing research upon which they hoped official policies would be built. Participant 1's research focused on best practices for SEO on the main website, whereas Participant 2's research assessed her library's overall online presence, including both the library website and the OPAC.

This lack of policies may present a barrier to visibility on search engines for both libraries. 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.

Structures

Both participants described structural barriers to search engine visibility within their organizations. The departmental structure of Participant 1's library means that web development and SEO happen in a different department (and separate physical branch of the organization) from the development of the OPAC. The visibility of OPAC records in the broader online world does not seem to fall into either department's responsibilities, making it difficult for the organization to address the issue, or even be aware that an issue exists.

Participant 2's organization has mitigated that structural divide by creating a team with members from three different departments to consider the issue. However, Participant 2 identified another structural issue affecting her library's ability to control its online presence: vendors. Although her organization is interested in submitting sitemaps to search engines and improving the metadata in its OPAC records, these initiatives cannot simply be carried out by library employees. The library's OPAC is provided by a vendor, who retains all webmaster privileges for the OPAC, as well as control over how resources appear online. Vendors save libraries time and money by taking care of OPAC records, but one consequence is that, as Participant 2 explained, "we don't have control over them submitting sitemaps, them dealing with robots.txt files, them adding microdata to the catalogue". This division of labour and control between libraries and vendors presents a barrier to immediate adoption of SEO practices by libraries for OPAC content. Instead, once best practices are determined, a dialogue must be opened with vendors in order to implement them. This may be seriously limiting libraries' capacity for experimentation in this area.

Technology

Both participants perceived technological barriers to making OPAC content visible in search engine results. Participant 1 expressed a general sense that attempting to crawl a library OPAC would be impractical for an indexing robot, due to the quantity and changeability of the records:

> What Google does is they just send out a Google bot and it just goes throughout the internet indexing pages over and over again. If it decided to index our catalogue, it would have to go to like a million different pages, catalogue each library holding, and the problem with that is, to stay current they would have to Google bot our catalogue every day, or every few days, because I don't know, two titles can change, right, in our library holding, but Google bot doesn't know the changes, it has to do the entire catalogue just to make those two changes.

Participant 2 also perceived this as a barrier, noting that while initiatives like Libhub attempt to revolutionize library record-keeping, libraries continue to struggle with the basic building-blocks of an indexable catalogue: stable, 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. Library 2'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 Participant 2, is that changeability identified by Participant 1. 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.

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".

Another barrier to true visibility relates to geolocation. Many search engine users expect their search results to take their location into account. A search for "pizza" is expected to produce local pizza restaurants, rather than global results. While one might expect the same principle to apply to a book search, Participant 2 reported difficulty being visible in local results, even when items were 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 (search engine results pages). 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.

This difficulty shows that although Library 2 has certainly made progress, true visibility on search engines is an elusive goal.

Because of these continuing issues, Participant 2 suggested that libraries must work on two different levels at the same time. On one level, libraries must work with initiatives like Libhub and Schema Bib Extend to create new ways of sharing library data with the rest of the web. On another level, libraries must work on "dealing with what we have going on right now, which is MARC and how MARC is then translated onto our Bib or Item pages on our OPAC".

Results: website indexing

A series of site-specific searches in Google and Bing revealed a general picture of each library's actual online visibility (as shown in Table I). Library 1 had less than five thousand OPAC pages indexed in each search engine. In addition, although individual pages were being indexed, the site's robots.txt page, which controls the actions of search engine crawlers, was not allowing the search engine to display any details about each page. This caused each search result to look the same, providing only the name of the library's OPAC (see Figure 1). Clicking on search results led to records for

Search Engine	Number of OPAC pages indexed Library 1	Library 2
Google	1,140	238,000
Bing	4,870	37,000

Table I.
Total indexed pages

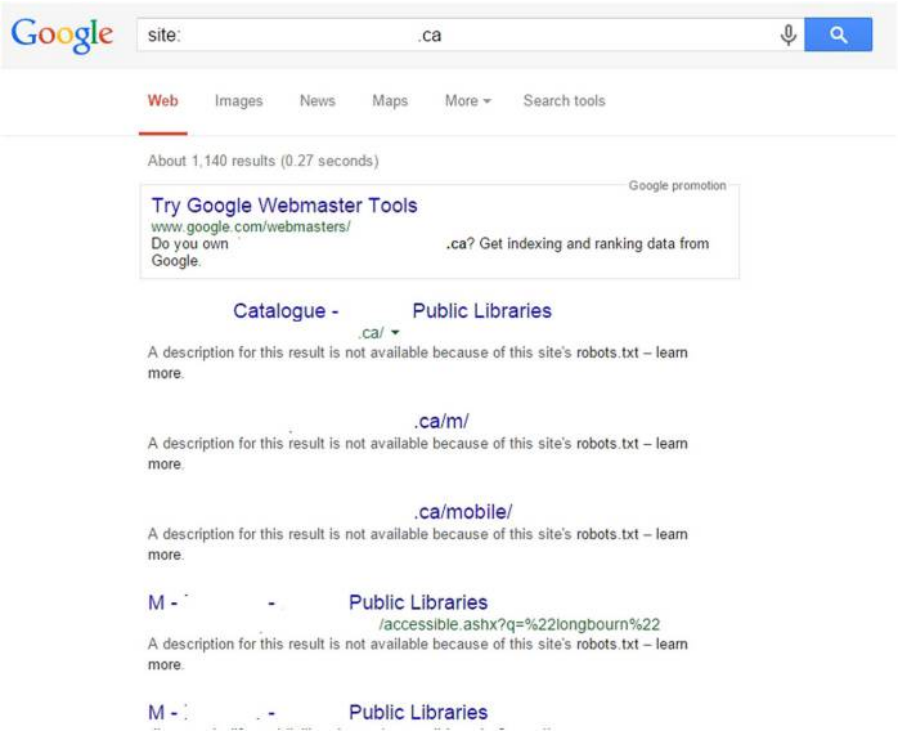

Figure 1.
Library 1's search results

individual resources, but the information indexed by search engines did not provide any indication as to where each link would lead. This lack of metadata essentially caused even the resources which were being indexed to be invisible. As a result, a more accurate total of the pages indexed in a usable way for Library 1 would be 0.

In contrast, Library 2 had tens of thousands of pages indexed in Bing, and hundreds of thousands in Google. Not only did Library 2 have more records indexed, but each search result included the item's title and author, a brief description of the item, the library's name and the name of their OPAC vendor (see Figure 2). This level of detail allowed results to be discovered in searches for specific items, and easily differentiated from each other. However, these results represented only a small fraction of the actual number of resources available through its OPAC. Participant 2 indicated that the total number of records listed in her library's OPAC fluctuate significantly, but was able to offer an approximate estimate of 575,799 records. Based on this number, the number of pages listed on Google represented about 41 per cent of total OPAC records, while Bing had indexed only about 6 per cent. In addition, as Participant 2 noted, the actual visibility of these results in an average search is still in question.

Results: website traffic

Being indexed is only the first of many steps in becoming visible on search engines. No further progress is possible without taking that first step, but having thousands of pages indexed does not necessarily translate immediately into thousands of patrons being referred to one's OPAC by search engines. This reality was evident in an analysis of website traffic statistics recorded by both libraries.

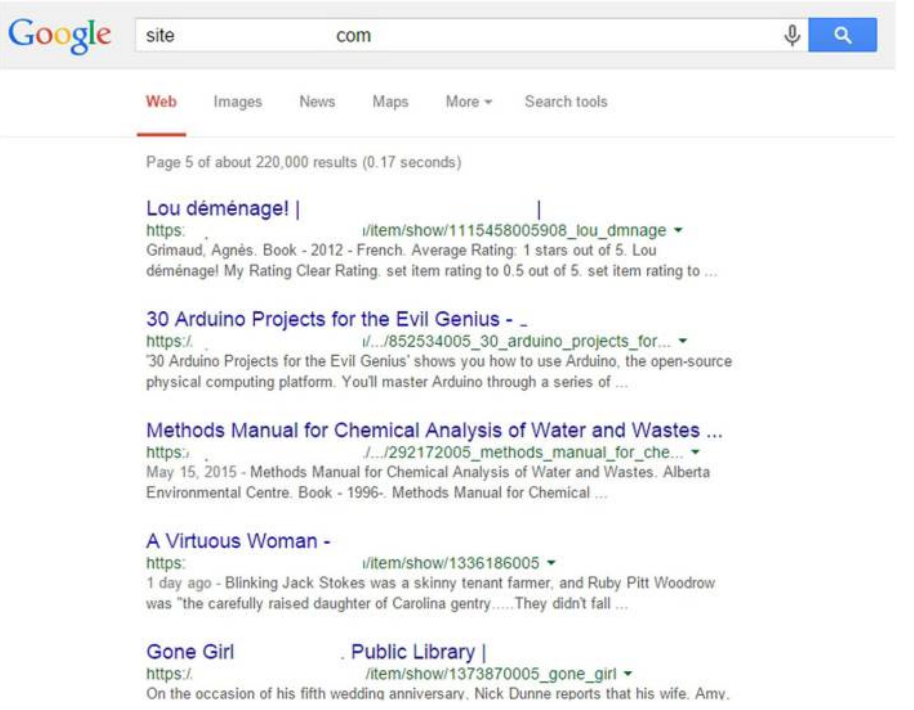

Figure 2.
Library 2's search results

Although specific data were unavailable regarding the source of visits for Library 1's website and OPAC, data were available for the most common landing pages, or first points of contact for visitors from April 2014 to January 2015. In this table the "Other" category encompasses all of the less popular landing pages put together, none of which accounted for a large enough share of traffic to be mentioned here (Table II).

This data showed that only 7.78 per cent of visits began with the OPAC. In contrast, the front page of the library's main website, including the mobile version, was the point of entry for 67.23 per cent of visitors. This is interesting, in light of Participant 1's statement when discussing website traffic data:

> One of the most important things that that has taught us as web developers is that our catalogue is our most important, most popular. I think it's like 63% of the traffic of our website is the catalogue itself.

This would suggest that although the OPAC accounted for the majority of traffic overall, most users first navigated to the library's main website as an entry point to the OPAC. This aligned with Participant 1's understanding of visitors' site usage: "they do go to the website and then the OPAC".

While Library 2 had succeeded in having a fair portion of its OPAC indexed, this progress was not yet reflected in the way users were finding OPAC resources. Website traffic statistics recorded for the year of 2014 showed that the majority (72 per cent) of OPAC visitors reached the site through referral from another site. The next most common source of visitors was direct access (i.e. bookmarks or URLs typed directly into browsers), at 26 per cent. Only 2 per cent of visitors found the OPAC using search engines. In contrast, the main site received 30 per cent of its visitors through organic search (Table III).

Further investigation of the referrals which accounted for most of the OPAC's traffic showed that the vast majority (97 per cent) of them were coming directly from the library's main website. The library's main site, which was much more visible on search engines than its OPAC, was the main entry point to the OPAC itself.

Most popular landing pages

URL	Percentage of sessions
Main Site	62.32
OPAC	7.78
Main site (mobile version)	4.91
Main site index	2.03
Other	22.96

Table II. Library 1's most popular landing pages

Source	OPAC (%)	Main Site (%)
Referral	72	14
Direct	26	54
Organic search	2	30
Social	0	1
E-mail	0	1

Table III. Source of visitors for Library 2's OPAC and main site

Discussion

The above results provide a snapshot of two libraries at different stages in their relationship with search engines. Even in this small, exploratory study, a range of attitudes can be seen. Both sides of the debate outlined in the above literature review are reflected in this study's results: from Participant 1's suspicion, which mirrors the negative perceptions expressed by authors such as Norris (2006), Herring (2005), and Gorman (2006), to Participant 2's enthusiasm, which corresponds to the views of authors such as Arlitsch (2014) and Thurow (2015). Participant 1 did not see being indexed by search engines as a goal, and described an institution that is apparently unaware of the possibility. Participant 2 was part of an enthusiastic team working to make her library more visible on search engines. This difference in attitudes was the most marked contrast between the two participants and their libraries.

As noted, Participant 1 is not alone in being sceptical of search engines' motives (e.g. Granka, 2010; Goldman, 2008; Hinman, 2008). Many are troubled by the idea of the world's information being organized by a private company, whose underlying criteria for selection and relevance ranking are hidden behind a "corporately-justified veil of secrecy" (Hinman, 2008, p. 73). From this perspective, Participant 1's reluctance to participate in a possibly flawed or corrupt system seems to support libraries' commitment to non-partisan and transparent public information access. However, it could be argued that most straightforward way to influence the scope and quality of information available through search engines is to participate in them. If information professionals believe that the information resources within their purview have value to the public, making those resources easily accessible through the most popular public information source should be a pressing priority. Not only is this aligned with libraries' mandate to make information as accessible as possible; it is also in their best interest. The more visible libraries are online, the better they can reach the public they serve. Without that visibility, users may not be aware of libraries as options for deeper, more complete information access.

While their attitudes regarding search engines differed, both participants reported similar barriers to visibility. Both libraries lacked written policy relating to search engines, and both were in the process of creating new policies to address that gap. In both cases, the internal structure of the organizations made it difficult to address the issue of search engine visibility. This supports Arlitsch's *et al.* (2013) argument that including SEO in institutional structures and policies is a crucial part of addressing search engine visibility. In addition, both participants perceived continuing technological barriers to being visible, and the quantitative data confirms those concerns. Some of the observed barriers, such as lack of SEO policies, were in the process of being addressed. Others, such as the attitudes of suspicion and indifference expressed by Participant 1, actively reduced the library's chances of addressing, or even being aware of, any other barriers.

It is interesting to note, however, that some of the perceived technological barriers noted by participants are not necessarily seen as barriers in the broader online world. For instance, although Participant 1 was correct in noting the large volume of data contained in OPACs, the number of library records to be crawled is less daunting when compared to the vast volumes of data already being crawled every day by search engines. Similarly, while maintaining permanent URLs is difficult, doing so is not generally seen as a necessary prerequisite for being crawled by search engines. Content providers simply expect search engines to crawl their pages frequently in order to correct any URL changes which have taken place. The changeability of the internet is widely acknowledged, and an occasional 404 error is par for the course. Participant 2's desire to conquer this problem before prioritizing site indexing lends credence to

Arlitsch's (2014) suggestion that librarians' conscientious attachment to high data standards may be hindering their ability to mesh with the rest of the internet.

Although some of these perceived technological barriers might be dismissed by a seasoned webmaster, the question of how to reconcile libraries' legacy systems, such as MARC, with current online conventions is still in the process of being answered. It is now technically possible for search engines to index many OPAC resources, but helping crawlers do so in a meaningful and useful way continues to be a challenge. Nevertheless, Participant 2 expressed her hope that libraries can work on several levels simultaneously in order to deal with immediate and more long-term online visibility goals:

> I think that's the nice thing though about now, it doesn't have to be just one thing. I mean back in the day we chose MARC, right, we stuck with MARC. It doesn't just have to be just MARC now, right? We can work in MARC and have our records converted into BIBFRAME which will also be marked up with microdata from schema.org. It doesn't have to be just one thing. We can work in multiple layers until we figure the whole thing out.

This attitude offers hope that libraries can feel their way forward into new terrain, addressing immediate visibility needs at the same time as they build solid, long-term structures. A many-layered approach is also advocated by both linked data initiatives discussed in the literature review; Libhub and Schema Bib Extend hope to work together, eventually, to provide a compromise between the granularity needed by libraries and the simpler data needed by search engines (Godby and Denenberg, 2015). In this view, libraries may not, as Arlitsch (2014) suggests, have to give up entirely on their beloved data standards; they must simply work to integrate those standards into a larger overall approach which offers a higher level of complexity for cataloguing librarians, and a lower level of complexity for search engines and patrons.

The quantitative data assessed in this study reflected both the similarities, and the major difference between the two libraries. The enthusiasm felt by Participant 2 was reflected by the data in terms of total OPAC pages indexed by search engines. By that criteria, Library 2 appeared to be much more visible than Library 1. However, the barriers shared by both libraries were also reflected in the quantitative data. While in both cases the OPAC was a very popular part of the library's online presence, most patrons of both libraries reached library resources by first navigating to the main website and then following a link to the OPAC. Library 2 may have had many thousands of OPAC resources indexed by search engines, but that progress did not, as yet, appear to be having a substantial effect on the way patrons were actually finding the library's resources online. In addition, Participant 2 noted that although 2 per cent of the OPAC's traffic came from search engines, the majority of the search terms being used were still library-specific (e.g. "x public library search"), rather than being searches for specific items. This implies that even that 2 per cent of users were already aware of the library, rather than stumbling upon it while searching for a resource. This is in contrast to the results reported by Boston (2005) and Breeding (2006), who both found that allowing a collection to be indexed caused a marked increase in usage. It is possible this difference was due to the difference in type of collection (both Boston and Breeding were working with digital collections, rather than local physical resources), or perhaps the format of Library 2's currently indexed content was less amenable to search engines' relevancy ranking algorithms. Without further research, any proposed explanations remain speculative.

This data suggests that Participant 2's positive attitudes towards search engine visibility had not changed the overall pattern of website use experienced by the library.

In the case of both libraries, the number of patrons who actually found each library's OPAC content through search engines was miniscule to non-existent.

It is important to note that both libraries were in a period of transition. Both participants were in the process of researching best practices and assessing their libraries' position. While Participant 2 and her colleagues were interested in SEO, they had yet to implement any related policies or practices. Their initiative was still in the exploratory phase. It is possible that quantitative data collected in a year or two would tell a very different story. For now, the data shows that both libraries have a long way to go in order to be truly visible through search engines.

Conclusion

Participant 2 summed up the issue libraries face: "You have to know where the front door is to get in". In other words, you have to start at a library's main site in order to access or even see the resources offered by that library. Many faithful library patrons take that route, but how many potential patrons never think to search specifically for the library website when they have an information need? In the case of both libraries in this study, the chances of an online searcher stumbling upon a library resource without consciously seeking the library were slim.

This study has explored a complex combination of barriers that stood between each library and visibility on search engines. Some of those barriers, such as lack of SEO policies, were in the process of being addressed. Others, such as the attitudes of suspicion and indifference expressed by Participant 1, actively inhibited the library's ability to address, or even be aware of, any other barriers. Even in this small, exploratory study, a range of attitudes could be seen: from Participant 1's suspicion, to the indifference he described in his colleagues, to Participant 2's enthusiasm. However, despite these different attitudes, similar patterns were evident in each library's usage statistics. Library 2 may have had many resources indexed by search engines, but that progress did not, as yet, appear to have had a noticeable effect on the way patrons were actually finding the library's resources online.

It will be interesting to observe the progress of each library in the coming years. Both institutions were in a period of transition when it comes to their online presence. It seems likely that many of the factors observed in this study will change significantly as each library completes its exploratory research, builds policy, and begins to tackle the observed technological and structural barriers. Future research documenting and assessing this progress would be valuable.

Further research is also necessary to determine if the factors identified in this study are more widely applicable within the public library community. Examining and comparing the challenges facing public libraries can offer useful information not only to researchers, but also to libraries themselves. Research into this rapidly evolving issue has the potential to help libraries compare notes, learn from each-others' mistakes and triumphs, and better understand the barriers and the opportunities inherent in online visibility initiatives.

References

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

Arlitsch, K., O'Brien, P. and Rossmann, B. (2013), "Managing search engine optimization: an introduction for library administrators", *Journal of Library Administration*, Vol. 53 Nos 2-3, pp. 177-188, doi: 10.1080/01930826.2013.853499.

Bazeley, P. (2013), *Qualitative Data Analysis: Practical Strategies*, Sage Publications, Los Angeles, CA.

Blandford, A. (2015), "Google, public libraries, and the deep web", *Dalhousie Journal of Interdisciplinary Management*, Vol. 11 No. 1, pp. 1-18, doi: 10.5931/djim.v11.1.5525.

Boston, T. (2005), "Exposing the deep web to increase access to library collections", paper presented at The Eleventh Australasian World Wide Web Conference, Gold Coast, available at: http://ausweb.scu.edu.au/aw05/papers/refereed/boston/paper.html (accessed 20 January 2015).

Breeding, M. (2006), "Systems librarian: how we funnelled searchers from Google to our collections by catering to web crawlers", *Computers in Libraries*, Vol. 26 No. 4, pp. 22-25.

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

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

Connaway, L.S. and Randall, K.M. (2013), "Why the internet is more attractive than the library", *The Serials Librarian*, Vol. 64 Nos 1-4, pp. 41-56, doi: 10.1080/0361526X.2013.761053, available at: www.tandfonline.com/toc/wser20/64/1-4

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

Egger-Sider, F. and Devine, J. (2005), "Google, the invisible web, and librarians", *Internet Reference Services Quarterly*, Vol. 10 Nos 3-4, pp. 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*, Vol. 140 No. 5, p. 18, available at: lj.libraryjournal.com/2015/02/technology/ending-the-invisible-library-linked-data (accessed 3 July 2015).

Enis, M. (2015b), "OCLC works toward linked data environment – ALA midwinter 2015", *Library Journal*, Vol. 140 No. 5, p. 18, available at: lj.libraryjournal.com/2015/02/technology/oclc-works-toward-linked-data-environment-ala-midwinter-2015 (accessed 3 July 2015).

Godby, C.J. and Denenberg, R. (2015), *Common Ground: Exploring Compatibilities Between the Linked Data Models of the Library of CONGRESS and OCLC*, Library of Congress and OCLC Research, Dublin, OH, available at: www.oclc.org/content/dam/research/publications/2015/oclcresearch-loc-linked-data-2015.pdf (accessed 3 July 2015).

Goldman, E. (2008), "Search engine bias and the demise of search engine utopianism", in Spink, A. and Zimmer, M. (Eds), *Web Search: Multidisciplinary Perspectives*, Springer, Berlin, pp. 121-133.

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

Granka, L.A. (2010), "The politics of search: a decade retrospective", *Information Society*, Vol. 26 No. 5, pp. 364-374, doi: 10.1080/01972243.2010.511560.

Halavais, A. (2013), *Search Engine Society*, Wiley, Hoboken, NJ.

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

Hinman, L.M. (2008), "Searching ethics: the role of search engines in the construction and distribution of knowledge", in Spink, A. and Zimmer, M. (Eds), *Web Search: Multidisciplinary Perspectives*, Springer, Berlin, pp. 67-76.

Lee, S., Jang, W., Lee, E. and Oh, S. (2016), "Search engine optimization: a case study using the bibliographies of LG Science Land in Korea", *Library Hi Tech*, Vol. 34 No. 2, pp. 197-206, available at: http://dx.doi.org/10.1108/LHT-02-2016-0014.

Norris, B. (2006), "Google: its impact on the library", *Library Hi Tech News*, Vol. 23 No. 9, pp. 9-11, doi: 10.1108/07419050610725012.

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

Östlund, U., Kidd, L., Wengström, Y. and Rowa-Dewar, N. (2011), "Combining qualitative and quantitative research within mixed method research designs: a methodological review", *International Journal of Nursing Studies*, Vol. 48 No. 3, pp. 369-383, doi: 10.1016/j.ijnurstu.2010.10.005.

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

Ryan, G.W. and Bernard, H.R. (2003), "Techniques to identify themes", *Field Methods*, Vol. 15 No. 1, pp. 85-109.

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

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

Thurow, S. (2015), "To optimize search, optimize the searcher", *Online Searcher*, Vol. 39 No. 4, pp. 44-48.

Uyar, A. and Aliyu, F.M. (2015), "Evaluating search features of Google Knowledge Graph and Bing Satori: entity types, list searches and query interfaces", *Online Information Review*, Vol. 39 No. 2, pp. 197-213, doi: 10.1108/OIR-10-2014-0257.

Zepheira, L.L.C. (2014a), "Libhub frequently asked and common questions", available at: www.libhub.org/faq/ (accessed 15 March 2015).

Zepheira, L.L.C. (2014b), "Join the movement – take the Libhub initiative pledge", available at: www.libhub.org (accessed 15 March 2015).

Corresponding author
Zoe Dickinson can be contacted at: zz979758@dal.ca