

Users and Uses of a Global Union Catalog: A Mixed-Methods Study of WorldCat.org

Simon Wakeling and Paul Clough

Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK. E-mail: s.wakeling@sheffield.ac.uk, p.d.clough@sheffield.ac.uk

Lynn Silipigni Connaway

Research, OCLC Online Computer Library Center, Inc., 6565 Kilgour Pl, Dublin, OH 43017, USA. E-mail: connawal@oclc.org

Barbara Sen

Information School, University of Sheffield, Regent Court, 211 Portobello, Sheffield, S1 4DP, UK. E-mail: b.a.sen@sheffield.ac.uk

David Tomás

Department of Software and Computing Systems, University of Alicante, Apartado 99. 03080 Alicante, Spain. E-mail: dtomas@dlsi.ua.es

This paper presents the first large-scale investigation of the users and uses of WorldCat.org, the world's largest bibliographic database and global union catalog. Using a mixed-methods approach involving focus group interviews with 120 participants, an online survey with 2,918 responses, and an analysis of transaction logs of approximately 15 million sessions from WorldCat.org, the study provides a new understanding of the context for global union catalog use. We find that WorldCat.org is accessed by a diverse population, with the three primary user groups being librarians, students, and academics. Use of the system is found to fall within three broad types of work-task (professional, academic, and leisure), and we also present an emergent taxonomy of search tasks that encompass known-item, unknown-item, and institutional information searches. Our results support the notion that union catalogs are primarily used for known-item searches, although the volume of traffic to WorldCat.org means that unknown-item searches nonetheless represent an estimated 250,000 sessions per month. Search engine referrals account for almost half of all traffic, but although WorldCat.org

effectively connects users referred from institutional library catalogs to other libraries holding a sought item, users arriving from a search engine are less likely to connect to a library.

Introduction

Sustaining the relevance and usefulness of library services in a networked age continues to challenge both professional and research communities. The recognition that institutional systems often fail to meet users' expectations has led to a paradigm shift toward systems that better facilitate single-point resource discovery and evaluation. At an institutional level this has meant the development and implementation of next-generation catalogs and discovery layers, offering users a single point of access to previously disparate collections and databases, and supplementing basic search functionality and collection metadata with additional features and content, such as faceted browsing, tags, reviews, and recommendations (Ballard & Blaine, 2011; Breeding, 2010). However, despite these attempts to realign library services with users' expectations, numerous studies still show the web as the starting point for many information seekers (Connaway, 2007; Kitalong, Hoepfner, & Scharf, 2008; Little, 2012). Integrating institutional library collections with popular web-scale discovery tools, particularly search engines, remains an ongoing and important challenge.

Received September 14, 2015; revised December 16, 2015; accepted December 17, 2015

© 2016 The Authors. Journal of the Association for Information Science and Technology published by Wiley Periodicals, Inc. on behalf of Association for Information Science and Technology. • Published online 0 Month 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23708

A potential means of tackling this issue lies in the utilization of the *union catalog*: “a catalogue that contains not only a listing of bibliographic records from more than one library, but also locations to identify holdings of the contributing libraries” (Feather & Sturges, 2003, p. 451). As “next-generation” catalogs have unified collections at the micro (institutional) level, union catalogs do so at a macro level, be it consortia, national, or global. Although the scope and purpose of union catalogs vary dramatically, a number of thinkers have highlighted the potential for large aggregated catalogs to be indexed by search engines, thereby facilitating discovery of, access to, and maximizing the value of disparate library collections. Dempsey (2006), for example, has observed that to match supply and demand, libraries “need new services that operate at the network level, above the level of individual libraries.” Similarly, Teets and Goldner (2013, p. 436) argue that libraries “need to expose the vast wealth of library collections data produced in the last 50 years beyond the library community.” Union catalogs, as preexisting aggregations of multiple library holdings, clearly have a role to play in realizing this vision. However, for union catalogs to fulfil their potential and meet users’ needs and expectations a clear understanding of the likely information needs and tasks users engage in as they access information is required (Allen, 1996). Despite the numerous user studies that exist for library catalogs, very little attention has been paid to union catalogs, particularly those at the global level.

In this paper we investigate the users and uses of WorldCat.org. Operated by OCLC, the global library collective, WorldCat is the largest bibliographic database in the world, with more than 300 million bibliographic records and more than 2 billion holdings from more than 70,000 libraries across the globe (OCLC, 2015). Since 2003, its records have been indexed by search engines and linked from Google Books. The catalog is directly accessible via a web interface (<http://www.worldcat.org>), which offers a range of standard library catalog discovery features, and as well as standard bibliographic data provides a range of supplementary information about items. This includes user-generated reviews and ratings (both added directly to WorldCat.org and imported from third parties, such as Goodreads.com), and links to online retailers selling the item. The system also offers a “Find a copy in the library” function, which links users to libraries geographically close to them that hold the item being viewed.

WorldCat.org has been the subject of research in a number of areas, including benchmarking for collection development (Perrault, 2002), analysis of holdings coverage (Bernstein, 2006), the identification of last copies (Connaway, O’Neill, & Prabha, 2006), and as a point of comparison to Google Books (Chen, 2012; Lavoie, Connaway, & Dempsey, 2005). However, there has been limited investigation of WorldCat.org usage and the information searching behavior of its users (Calhoun, Cantrell, Gallagher, & Cellantani, 2009; Nilges, 2006). This paper describes the largest study to date that seeks to investigate

the users and uses of WorldCat.org using a mixed-methods approach. Results from focus groups involving 120 participants, an online survey with 2,918 responses, and analysis of transaction logs involving around 15 million sessions are integrated to provide a more holistic view of WorldCat.org usage. The contributions of this paper are threefold. First, we provide an in-depth study of the users of WorldCat.org and their uses of the system; second, we present a categorization of work and search tasks from WorldCat.org that are applicable to union catalogs more widely; third, we demonstrate how multiple methods can be utilized for studying union catalogs, including the integration of data to form a holistic view of information-searching behavior.

The study seeks to address three research questions:

- [RQ1] What are the demographics (age, gender, location, and occupation) of WorldCat.org users?
- [RQ2] Where are users of WorldCat.org being referred from?
- [RQ3] For what purposes are users accessing WorldCat.org?

There are two principle benefits of this research. A common feature of models of information-seeking behavior is the recognition that the information-seeking process is essentially “the advance from uncertainty to certainty” (Wilson, 1999, p. 265). For those responsible for developing systems that support information seeking, that outcome is related to “the perceived need for information that leads to someone using an information retrieval system” (Shneiderman, Byrd, & Croft, 1997, Appendix 1). It follows, therefore, that for researchers seeking to improve system performance and user experience there is clear value in better understanding and classifying users’ needs (Gisbergen, Most, & Aelen, 2007; Rose & Levinson, 2004). Therefore, we expect that the results of this study will influence potential improvements to WorldCat.org. Second, we suggest that a better understanding of how WorldCat.org is currently used has the potential to inform the development of other union catalogs, and in particular to contribute to the ongoing debate concerning the methods and value of exposing library collections to wider audiences.

The methods described here also generated a rich data set relating to user search behavior and modes of interaction with WorldCat.org. Analysis of these data and a discussion of the implications for union catalog system design will be published shortly in a separate paper.

The remainder of the paper is structured as follows. The following section provides a review of the literature relating to union catalogs, and classifying user’s reasons for accessing library catalogs. Following this, we describe the multi-phase mixed-methods methodology used to collect and analyze data. This is followed by the integrated presentation of results relating to each research question, and a discussion of the users and uses of WorldCat.org and the impact of the findings more generally.

Literature Review

In this section, we discuss two areas of relevant literature. We first examine work related to union catalogs, and in particular studies that take a user-oriented approach. We then discuss ways in which the uses of library catalogs have been classified.

Union Catalogs

Broadly speaking, the literature on union catalogs can be divided into the conceptual and practical. From the conceptual perspective, some authors maintain that the traditional role of the union catalog is primarily a driver for interlibrary loan and resource sharing (Gorman, 2007; Hider, 2004). Others, however, see potential for union catalogs to play a broader role in the new information landscape. Lass & Quandt (2004) argue that the traditional uses of union catalogs (shared cataloguing, quality control, interlibrary loan) have been expanded to include the possibility of online search and text delivery with a single point of access. This intersection with web services is best examined by Gradmann (2004), who notes that although the exposure of union catalogs on the web is essential, the fundamental differences in approach between library and web systems must be acknowledged. In practice this means recognizing that “library-based information systems are based on the idea of mediated access, whereas the original principle of web-based systems is one of direct, instant access” (Gradmann, 2004, p. 77).

From a practical perspective, a number of authors have discussed information architecture issues relating to union catalogs, particularly the relative strengths and weaknesses of distributed and centralized models (Cousins, 1999; Hider, 2004). In addition, there exist a number of case studies detailing the technical and organization requirements behind establishing new or improved union catalogs (Alam & Pandey, 2012; Boston, Rajapatirana, & Missingham, 2009; Burnhill & Law, 2005; Larsen, 2007; Mittal, 2011). A further subset of the union catalog literature describes more user-oriented studies. Hartley and Booth (2006) present a study investigating how individuals use and view union catalogs, comparing COPAC (a union catalog of more than 70 UK and Irish University and Research libraries) with three UK regional union catalogs. Their methodology utilized observed search sessions, with volunteers completing predetermined tasks, interviews, and focus groups. As the authors note, the search scenarios developed for the research were based on “search types which experience had suggested. . . are put to union catalogs” (2006, p. 13), and the study therefore does not present empirical data relating to how and why union catalogs are used in the real world. Further work on COPAC is reported by Craven, Johnson, and Butters (2010), who gather data from 12 postgraduate students and academic staff using focus-groups, interviews, and controlled search tasks to examine the usability of the catalog.

Goodale and Clough (2012) take a more holistic approach in their user evaluation of the SEARCH25 system

(<http://www.search25.ac.uk>), a prototype successor to InforM25, the union catalog of more than 60 members of Academic Libraries in the southeast of England. Their study includes a survey of users, as well as log file analyses and focus group sessions. The survey reveals the most common tasks for which users frequently use the system relate to known-item searches, with 85% of respondents doing this often or very often. Discovery tasks, such as searching by subject, are less popular, although more than half of all users (59%) still regularly conduct these searches. The survey also indicated that users most valued SEARCH25 for its item coverage, seeing the system as a potential “one-stop-shop.” Analysis of a sample of the search logs revealed the average (mean) number of actions per session to be 3.8, with a majority of sessions (53.8%) consisting of just one action, and 85% of sessions consisting of five actions or fewer. The report also highlights some typical use scenarios, gleaned from focus group sessions with users of the system. Two of the scenarios represent a librarian using the system, either undertaking cataloguing and/or assisting a patron find an item at a reference desk, whereas the other two involve a student or researcher finding a comprehensive and diverse range of material on a topic, and determining which libraries hold certain collections.

Some prior research has examined the users and uses of WorldCat.org itself. For example, Nilges (2006) reported usage patterns from the initial integration of WorldCat.org with search engines, focusing primarily on the access points to WorldCat.org and the types of search behavior exhibited by users. Based on a sample of log files, Nilges states that users are most likely to access WorldCat.org records via a two-to-four term keyword search, and that the WorldCat.org result was on average the sixth result displayed in Yahoo! Search results, although a substantial number of clicks were from results ranked outside the top 10, indicating that “WorldCat.org does serve a constituency of more determined researchers who tend to dig deeper into results sets” (Nilges, 2006, pp. 442–443). Users also were found to click on a “Find a Library” link about 4% to 6% of the time.

Calhoun et al. (2009) take a user-centered approach to the question of data quality in WorldCat.org, using end user focus groups, a pop-up browser survey for users accessing WorldCat.org, and a separate survey of librarians. The pop-up survey, which collected 11,151 total responses, showed librarians making up 32% of respondents, with postgraduate (15%) and undergraduate (13%) students making up a further 28%. Teachers and academics constitute 22%, with “Business Professional” and “Other” accounting for the remainder. Although the focus of the research was on existing data quality, and potential improvements to the system, the study distinguishes between two typical types of tasks that users undertake: (1) known-item, that is, accessing information about a particular preidentified item, and (2) discovery, that is, using the system to find and evaluate potentially useful items. Nilges acknowledges that these tasks make different demands on the system. Overall, the study notes that users of all types access WorldCat.org

purposefully, with librarians likely to be carrying out “work responsibilities,” and other users seeking resources to address some information need.

Perhaps the most notable aspect of the literature review conducted for this project is how little work has been done to identify who is using union catalogs and why they are using them. We also might conclude that although WorldCat.org is a fruitful source of research in a number of areas, there has yet to be research that focuses specifically on the makeup, needs, and behavior of its users.

Classifying Library Catalog Search Tasks

For the purposes of this paper we follow an existing interpretation of task-based activities based on Toms (2011). This identifies some work function to be the predicating condition of any information seeking, with work here understood in its broadest sense, relating not only to economic but any other “extrinsic benefit” (Toms, 2011, p. 44). Within this work context, an individual is likely to undertake tasks. Understanding tasks within a work context leads naturally to the conception of the term *work-task*, a term used by a number of authors to represent an overarching unit within which information-seeking activities are undertaken (Bystrom & Hansen, 2005; Vakkari, 2003). Work function can consist of any number of work-tasks, and each of the tasks may themselves consist of subtasks. One such subtask is the search-task, which represents the motivating external factors influencing user interaction with an information retrieval or support system.

Empirical studies examining the work and search tasks that motivate union catalog use are in short supply. However, a variety of attempts have been made to classify typical search tasks for which users engage institutional online library catalogs. Lewandowski (2010) maps catalog search tasks to Broder’s well-known taxonomy of web search, likening a known-item search to Broder’s (2002) *Navigational* classification, and a topic search to an *Informational* intent. For Lewandowski, the online catalog equivalent of the *Transactional* search is the search for sources, during which a user attempts to locate a source from which to continue their information seeking, for example, another database. Empirical studies of catalog use have developed alternative schemes. Hert (1996) based her analysis of user search tasks on observations of students interacting with the online catalog at Syracuse University. The various goals articulated by participants are reduced to four overarching types: a search for a specific known-item; a search for an unknown-item, that is, a single resource on a particular topic; a search for information about an item, for example, the start date of a journal; or a general search for information with no specific number or type of resource in mind. The notion of an unknown-item search is also found in Slone (2000), who attempted to categorize the search tasks of searchers using public library catalogs. Based on data collected from surveys, interviews, and observations of students, she identifies three key types of tasks: known-item, unknown-item, and

TABLE 1. Applicability of each research phase to research questions.

	Focus groups	Survey	Log analysis
RQ1 (User demographics)	-	X	X
RQ2 (Referrer)	X	-	X
RQ3 (Purpose of visit)	X	X	X

area. For Slone, the unknown-item category encompasses what other authors have termed *subject* or *topic searches*, but also incorporates search tasks that would only uncomfortably fit into the topic search category (e.g., searching for a single textbook). The area search relates to users who use the catalog to determine the area of the physical library items on a particular topic are held, and then continue their searching there.

The location of a known-item within the catalog is recognized as a core task within the classification schema described previously, and a number of studies of catalog use identify accessing a known-item as the most common search task in library catalogs (Larson, 1991; Yee & Layne, 1998). Yet as Lee, Renear, and Smith (2006) note, “most researchers articulate their own conceptual and operational definitions of a known-item search, making little effort to explicitly connect these to the general concept and rarely providing citations to sources or authorities” (p. 3). This study adapts Slone’s definition of a known-item search and defines it as an interaction with the system wherein the searcher is seeking to locate in the catalog the record of a specific item, about which some data are known. This is contrasted with an unknown-item search, which we define as an interaction with the system where the searcher is seeking to locate in the catalog one or more items that offer some potential utility, without knowing the specific items in advance.

Methodology

To effectively address the research questions, a pragmatic multiphase mixed-method methodology was devised. The design drew on a number of prior studies of library catalog use (e.g., Ballard & Blaine, 2011; Bertot et al., 2012; Craven et al., 2010), with research consisting of focus groups, an online pop-up survey, and analysis of WorldCat.org transaction logs. In addition to the benefits associated with individual quantitative and qualitative techniques, mixed-methods research offers the potential for complementary data sources to improve generalizability, provide stronger evidence for conclusions, and add insight and understanding (Johnson & Onwuegbuzie, 2004). Table 1 shows how results from each of the three phases related to the research questions.

The methods employed for each of the three phases are described next, with further details available in Wakeling (2015). Data collection was carried out between 2011 and 2013.

TABLE 2. Focus group interview participants by user group and location.

	Aus/NZ		UK		US		Total	
	No. sessions	No. participants						
All Librarians	6	23	3	20	2	20	11	63
Public Librarians	1	5	0	n/a	0	n/a	1	5
Academic Librarians	3	11	2	11	2	17	7	39
National Librarians	3	10	1	9	0	n/a	4	19
All Students	0	n/a	2	17	3	24	5	41
Undergraduate Students	0	n/a	1	9	2	16	3	25
Graduate Students	0	n/a	1	8	1	8	2	16
Booksellers	0	n/a	1	9	0	n/a	1	9
Historians	2	4	2	3	0	n/a	4	7
Total	8	27	8	49	5	44	21	120

Phase 1: Focus Groups

Focus group interview research offers “a way of collecting qualitative data, which — essentially — involves engaging a small number of people in an informal group discussion ‘focused’ around a particular topic or set of issues” (Wilkinson, 2004, p. 177). Focus group interviews also constitute an established methodology within Library and Information Science (Connaway & Powell, 2010, pp. 173–174; Von Seggern & Young, 2003), and a number of previous studies have used the methodology to investigate the use of online catalogs (e.g., Berger & Hines, 1984; Connaway, Wilcox, & Searing, 1997). The intention of this phase of research was to gather qualitative data from users of WorldCat.org relating to their use of the system. The selection of groups to be targeted in the research was influenced by the survey results in Calhoun et al. (2009), and user personas created for internal use by OCLC. The user groups selected were librarians (public access and cataloguing; university and public), students (postgraduate and undergraduate), antiquarian booksellers, and academics (historians).

The questions asked during the focus group interview sessions were carefully designed to ensure that participants had the opportunity to address a broad range of issues and experiences with WorldCat.org. The research was conducted in three stages, each relating to a geographical location: Australia and New Zealand (March 21 to April 8, 2011), the United Kingdom (May 9 to 17, 2011), and the United States (October 25 to 27, 2011). Potential participants were identified using a purposive convenience and snowball sampling. The researchers drew on existing library contacts to assist with recruitment, except in the case of antiquarian booksellers, who were identified through their membership of professional bodies (the Australian & New Zealand Association of Antiquarian Booksellers, the Antiquarian Booksellers Association [UK], and the Antiquarian Booksellers’ Association of America). Although this approach was unsuccessful in Australia and the United States, we were able to recruit enough UK-based booksellers to conduct a focus group interview session. Student participants were compensated (£10 or \$20) for their involvement. The recruitment of historians proved most challenging. This academic discipline was selected as broadly representative of humanities schol-

ars, and historians were recognized by OCLC as key users of WorldCat.org, particularly for identifying and locating historical documents. However, despite exploring a number of avenues for recruiting academic historians, only seven eventually participated. Although this is a relatively small number, we note that focus group interviews are not generalizable and are used to familiarize one with specific areas of inquiry or to gather more in-depth information about specific areas of inquiry (Connaway & Powell, 2010). The focus group interviews for this research were therefore conducted to gather more information on specific types of WorldCat.org users, and were not intended to produce generalizable results. In total, 120 participants were interviewed during 21 sessions at 11 locations (Table 2).

Two researchers were present for each focus group interview session: one acting as moderator, the other as notetaker. The investigators alternated between roles. An audio recording of each session was made, and the notes from each session were augmented and clarified after a review of the audio recording. The results were analyzed using qualitative content analysis, following the process set out by Zhang and Wildemuth (2009). Both investigators closely examined the notes, highlighting all ideas and terms that related to participants’ engagement with WorldCat.org. These terms were then rationalized, merged as appropriate, and arranged into a hierarchical structure within five main categories: Work-Tasks, Search-Tasks, Strengths, Challenges/Difficulties, and Suggestions for Improvement. To test the code book, two researchers coded the same five randomly selected transcripts and compared results. After discussion, the code book was amended to reflect the final agreement on coding terms and organization, and the transcripts from all the focus group interview sessions were coded. Once all coding was complete, five sessions were randomly selected and coded by a colleague. The coding of these five sessions was compared for intercoder reliability using Cohen’s kappa coefficient and found to be at a level ($k = 0.85$) to indicate reliable coding (Yardley, 2008).

Phase 2: Survey

To achieve a more comprehensive understanding of users’ demographics and their reasons for accessing

TABLE 3. Classification of WorldCat.org referrers.

Referrer type	Description
Search engine	The referrer URL represents a web search engine. The final list comprised the following search engines: Google, Bing Yahoo, Yandex, Baidu, Sogou, Daum, Babylon, Delta-search, Ask.com, So.360.cn, Mysearchresults, Mywebsearch, and Searchmobileonline.
Library	The referrer URL represents a Library. This was captured using a regular expression to identify instances of a series of library related keywords within the referrer URL.
WorldCat.org home	The session starts directly at the WorldCat.org homepage (i.e., the first page loaded in the session is WorldCat.org, with no other referrer URL provided).
WorldCat.org other page	The referrer URL represents another WorldCat.org page. These might be part of the WorldCat.org identities service, or other pages with a worldcat.org url that do not constitute the catalog itself. It is also likely that a number of sessions assigned this classification will relate to lines from the log relating to a single IP address that have been split into two or more sessions. The second of these sessions would appear to have a WorldCat.org referrer url.
Citation service	The referrer URL represents a citation service (easybib.com, bibme.org, citefast.com, redlightgreen.com/org, or mendeley.com).
Goodreads.com	The referrer URL represents a GoodReads page.
Wikipedia.org	The referrer URL represents a Wikipedia page.
OCLC services	The referrer URL represents an OCLC page.
Other	The referrer URL is present in the logs, but does not map to any of the above categories.
Not specified	The referrer URL is absent or improperly formed in the logs. This most likely represents a web service that has blocked their referrer details.

WorldCat.org, invitations to complete an online survey were distributed via pop-ups on the WorldCat.org site. The survey questions were developed to cover two areas relevant to this study: (a) user demographics (gender, age, location, and occupation), and (b) purpose and reason for using WorldCat.org. The survey was pretested by a total of seven academics, students, and librarians and revised accordingly. The survey also sought to capture potential differences in behavior and intent between users accessing the site through the WorldCat.org homepage (by typing “worldcat.org” directly into a browser or using a bookmark), and those landing directly at detailed record pages (i.e., the page in the catalog relating to an individual holding), for example, by following a link from a search engine. Two identical questionnaires were therefore created in SurveyMonkey, and linked to pop-ups appearing either from the homepage or detail pages. The invitation to complete the survey was set to appear on every 100th record page accessed, and every 100th time the homepage was loaded, reducing the likelihood of a single user receiving multiple invitations.

The survey went live at 00:00 hours Eastern Standard Time on Thursday April 5, 2012. After a week, a review of completed surveys revealed an extremely low response rate from the WorldCat.org homepage. It was therefore decided that the invitation would be set to appear every time the homepage was loaded (rather than every 100th time) for the remainder of the survey period. Invitations at the record pages remained at 1/100. The survey ran with these invitation ratios until 00:00 hours Eastern Standard Time on Thursday April 19, 2012. A total of 980 responses were col-

lected from the WorldCat.org page survey and 2,669 from the record pages survey. Of these 3,649 responses, 731 were incomplete, leaving 2,918 completed surveys (894 from the .org page, 2,024 from record pages). Based on the traffic to WorldCat.org shown in the logs for October 2012, the response rate could be estimated at 1.6%. Although this is low for traditional survey instruments, it is not uncommon for online pop-up surveys to record response rates well below 5% (Ockuly, 2003).

Phase 3: Transaction Log Analysis

Transaction log analysis (TLA) describes the methodical and comprehensive investigation of queries and other actions executed by a user, and the resulting system response (Blecic et al., 1998; Phippen, Sheppard, & Furnell, 2004). Thus, TLA “can be conceptualized both as a form of system monitoring and as a way of observing, usually unobtrusively, human behavior” (Peters, 1993, p. 42). Log data for 2 months of WorldCat.org traffic (October 2012 and April 2013) were analyzed. Preparation of the log data included filtering out nonhuman traffic, such as web search engine crawlers, together with removal of sessions consisting of more than 100 queries (Jansen, 2006). The removal of robot traffic reduced the number of lines in the combined logs by more than half, from over 100 million to 56 million. Data preparation also included identification of user sessions. A time-based method using a 30-minute cutoff period was employed (Jones & Klinkner, 2008). A new session ID was therefore applied to logs originating from a single IP address if server transactions attributable to that IP address were separated by

Known-item

Timestamp	URL	Action Type
14:39:33	/ HTTP/1.1	Load WorldCat.org homepage
14:39:51	/search?qt=worldcat_org_all&q=little+star+john+hindqvist	Search WorldCat.org from homepage
14:40:27	/title/little-star-a-novel/oclc/778421881&referrer=brief_results	View an item after clicking in the search results page

Unknown-item

Timestamp	URL	Action Type
12:25:33	/ HTTP/1.1	Load WorldCat.org homepage
12:25:44	/search?qt=worldcat_org_all&q=quebec+CEGEP	Search WorldCat.org from homepage
12:25:55	/title/evaluating-programs-of-study-in-quebec-case-study-technical-education-programs-in-early-childhood-education-at-the-cegep-de-saint-jerome/oclc/38753252&referrer=brief_results	View an item after clicking in the search results page
12:25:57	/wcpa/servlet/org.oclc.lac.ui.ajax.ServiceServlet?serviceCommand=getCitation&oclcnum=38753252&style=MLA	Click on "Cite/Export"
12:26:07	/title/colleges-in-quebec-a-new-phase/oclc/25443005&referrer=brief_results	View an item after clicking in the search results page
12:26:08	/wcpa/servlet/org.oclc.lac.ui.ajax.ServiceServlet?serviceCommand=getCitation&oclcnum=25443005&style=MLA	Click on "Cite/Export"
12:26:14	/title/student-transitions-and-adjustments-in-canadian-post-secondary-education/oclc/377666503&referrer=brief_results	View an item after clicking in the search results page
12:26:15	/wcpa/servlet/org.oclc.lac.ui.ajax.ServiceServlet?serviceCommand=getCitation&oclcnum=377666503&style=MLA	Click on "Cite/Export"
12:26:16	/search?q=quebec+CEGEP&fq=&se=%24d&sd=desc&dblist=638&start=11&qt=n ext_page	Click "Next" link in the search results page
12:26:46	/title/analysis-of-quebec-cegep-graduates/oclc/61547347&referrer=brief_results	View an item after clicking in the search results page
12:26:48	/wcpa/servlet/org.oclc.lac.ui.ajax.ServiceServlet?serviceCommand=getCitation&oclcnum=61547347&style=MLA	Click on "Cite/Export"

FIG. 1. Examples of manually coded sessions.

at least 30 minutes of inaction. This cutoff period was used to assign unique session IDs to the full data set, and the final logs were found to constitute 15,799,727 sessions. Thus, WorldCat.org was found to support around 8 million sessions per month (October 2012 = 7,996,172; April 2013 = 7,803,555). However, initial analysis of the log files revealed that well over a third of these sessions (39.7%) consist of a single line in the log, which represents the loading of the landing WorldCat.org homepage or item-level record page.

To properly address RQ2, additional work was undertaken to classify the referrer type of any URL with more than 5,000 session instances in the log. This resulted in the 10 referrer categories shown in Table 3.

An additional stage of analysis involved the manual coding of three sets of sample sessions. The intention here was to infer the types of search task undertaken by users interacting with the system and to compare results for users who directly accessed the WorldCat.org homepage, users whose sessions originated from a search engine, and users arriving from a library referral. To capture sessions that involved some level of system interaction, 400 sample sessions that included at least one search action were extracted from the log for each of the three referrer types. This sample size was deemed sufficient based on precedents set in the literature relating to session classification (e.g., Broder, 2002; Jansen, Booth, & Spink, 2008). The main aim of the coding was to judge whether a session constituted a known-item or unknown-item search task or some combination of the two. The criteria used to determine the type of search task was

based on existing literature relating to known-item query formulation and detection. A number of authors have observed the frequency and effectiveness of known-item queries that combine author name and title (Kilgour, 2001; Slone, 2000). Kan and Poo (2005) highlight six characteristics of known-item queries that can aid identification. They posit that as well as being longer than topic search queries, known-item queries are more likely to contain determiners ("the," "a," etc.), proper nouns, mixed cases, advanced search operators, and object identifying keywords, such as "textbook" or "article." Because the analysis was conducted at a session rather than query level, it was also possible to identify occasions when the query terms precisely matched the title of an item subsequently viewed. The coding process itself involved essentially "replaying" each session by following the URLs contained in the log, where loading the page in a web browser to better understand the user's interactions was necessary. Examples of a sessions coded as known item and unknown-item are presented in Figure 1.

On completion of the coding, a random 20% of the raw sample sessions was extracted and recoded by another researcher with the same scheme with high intercoder agreement ($k = 0.89$).

Data Integration

The integration of data from the three research phases was guided by Bazeley and Kemp's (2012) metaphors for integrative analysis. Their work combines ideas from

TABLE 4. Geographical location of users: results from log analysis and survey.

	Log analysis		Pop-up survey	
	Country	% of total traffic	Country	% of survey responses
1	United States	44.8	United States	49.9
2	China	5.3	Canada	4.8
3	Canada	5.2	China	4.7
4	United Kingdom	3.7	Germany	3.7
5	Germany	3.2	United Kingdom	2.7
6	France	2.3	Australia	2.6
7	India	1.8	Brazil	2.0
8	Italy	1.7	India	1.9
9	Indonesia	1.7	Mexico	1.7
10	Spain	1.5	Italy	1.7
11	Netherlands	1.5	Netherlands	1.3
12	Mexico	1.3	France	1.0
13	Australia	1.3	Spain	0.8
14	Brazil	1.3	Belgium	0.7
15	Poland	1.2	Sweden	0.7
16	Japan	0.9	New Zealand	0.7
17	Malaysia	0.9	Russian Federation	0.7
18	Korea, Republic of	0.7	Switzerland	0.7
19	Russian Federation	0.7	South Africa	0.6
20	Singapore	0.7	Columbia	0.6

throughout the methodological literature into a set of approaches to data integration, which they express as metaphors. The result is a loose framework of methods, including completion (amalgamating findings into a unified whole), enhancement (mingling diverse but complementary findings), triangulation (cross-validation), exploration, and conversation (identifying and linking “sense strands”). A number of these techniques were used in the integration of data from the three research phases, with the process described in detail in Wakeling and Clough (2015).

Results

Geographical Coverage (RQ1)

The geographical spread of users found in the logs is very similar to that of survey respondents: 13 countries appear in both top-20 lists (ranked by number of sessions and respondents, respectively), and both lists show a large proportion of users coming from the United States (Table 4). That the spread of survey respondents appears so similar serves to partially validate the survey findings, at least to the extent that the respondent population can be shown to generally represent the geographic distribution of the total user population. As a whole, this study finds that WorldCat.org can justifiably be called a global service: More than 200 countries and territories are represented in the log data, and while North American traffic accounts for a large percentage of traffic, the long-tail of other countries represent around half of all users coming to the site.

Age, Gender, and Occupation (RQ1)

A slightly higher number of females than males completed the survey (female = 55.2%, $n = 1,611$; male = 44.8%,

$n = 1,307$). The age of participants was found to be high: 63.5% of respondents ($n = 1,852$) gave their age as 36 or older, 19% of respondents identified as being younger than 25, and 18% as being between 26 and 35 years. The age group 50 years and older was the best represented (39%, $n = 1,137$).

Survey respondents were asked to provide their occupation, with four options provided (undergraduate student, postgraduate student, librarian, and faculty/researcher), as well as an option to manually enter an alternative occupation, which were manually reviewed and grouped appropriately. A detailed breakdown of all occupations, including coding categories, can be found in Figure 2. Students (graduate and undergraduate) represent the largest single aggregate respondent group (35.9%, $n = 1,049$), whereas library staff (“Librarian” and “Other library staff”) account for a quarter of all respondents (25.1%, $n = 733$) and academic staff less than a fifth (17.3%, $n = 506$). Respondents identifying as “other” occupations make up the remainder (21.6%, $n = 630$).

Figure 3 shows a breakdown of the occupations of respondents from the 10 best-represented countries in the survey. It shows the United States and Canada as the only two countries to have a higher proportion of library staff respondents than students.

Referrals to WorldCat.org (RQ2)

Sessions originating from a search engine are by far the most common type found in the logs and represent almost half of all traffic to WorldCat.org (47.1%, see Table 5). Referrals from libraries account for a further 14.4% of sessions, whereas traffic from other WorldCat.org pages (6%), and sessions originating at the WorldCat.org homepage (5.3%) in total account for around 1 in 10 sessions in the

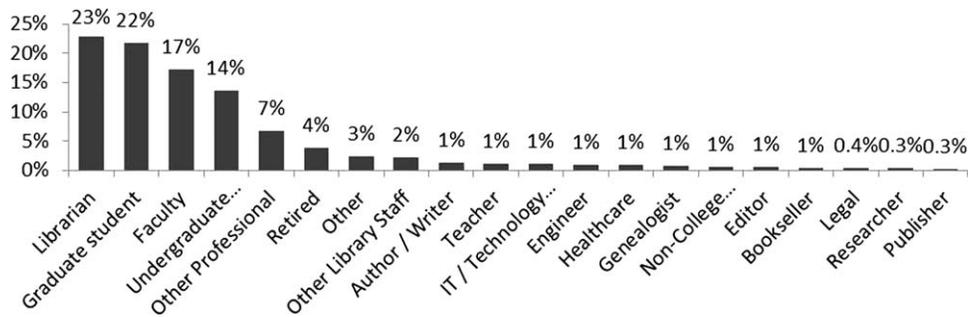


FIG. 2. Survey respondent occupations (n = 2,918).

logs. Although the overall proportion of sessions from citation services, GoodReads and Wikipedia are low, they still represent a significant number of visitors to WorldCat.org.

It was further possible to compare the distribution of referrer types originating from each country. Table 6 shows these distributions for the top 10 countries. The United States and Canada have the lowest proportion of their sessions originating from a search engine (29.8% and 30.2%, respectively), and the highest beginning directly at the WorldCat.org homepage (7.5% and 5.7%), likely reflecting increased awareness of the service in North America. Indeed, traffic from the US accounts for 87% of all sessions originating at the WorldCat.org homepage. For all other countries, the majority of sessions are referred from a search engine, with more than 70% of traffic from India, Italy, and Spain originating from that source. It should be noted that despite the proportion of US traffic originating from a search engine being relatively low, separately computing the distribution of search engine referred sessions between countries shows that more than a quarter (28.4%) of all such traffic originates in the United States.

Uses of WorldCat (RQ3)

Work-tasks. The focus group interview participants described three broad contexts for using WorldCat.org: professional, academic, and leisure. As might be expected, librarians and booksellers were the most likely to use

WorldCat.org for professional purposes. Several of the librarians who participated in the focus group interviews were cataloguers, and they spoke of using WorldCat.org as a means of establishing the bibliographic details of items they were required to catalog for their institution. Booksellers described using the system for similar reasons; in their case, adding book descriptions and metadata to their stock lists.

Survey respondents also were asked to classify their purpose for visiting the site as one of three options: educational, professional, or recreational. Only 13% (n = 378) of respondents had a recreational purpose for visiting the system, with the figures for key users groups for WorldCat.org—students (7.5%, n = 79), faculty (6.5%, n = 33), and library staff (6.0%, n = 44)—even lower. In contrast, 59.1% (n = 65) of retired respondents stated they were using the system for recreational reasons.

Librarians in the focus group interview sessions, particularly those working on reference desks or in other user-facing roles, spoke of how they used WorldCat.org to assist students and faculty with interlibrary loan (ILL) requests, whereas others had responsibility for collection development and acquisitions, explaining how they used WorldCat.org as a source of data to direct their strategic buying or collection optimization decisions. Booksellers mentioned using WorldCat.org to assist in the valuation of rare items (“to get a sense of relative rarity,” London Bookseller). One academic also described using the system during the process of developing and updating student reading lists. Finally, librarians involved in information literacy or other library training programs mentioned their use of the system during training and

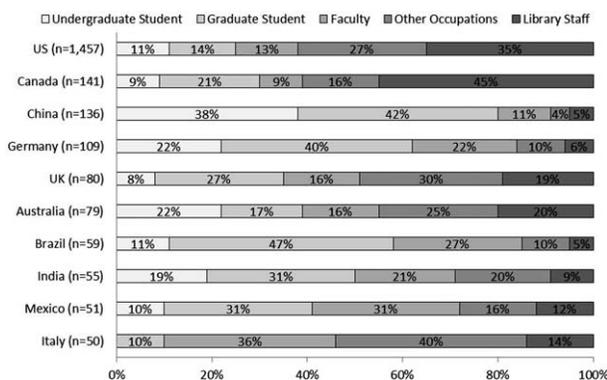


FIG. 3. Breakdown of survey respondents by occupation for top 10 countries (n = 2,918).

TABLE 5. Sessions originating from each referrer type based on the log data.

Referrer type	Sessions (n = 15,799,727)	% of total sessions
Search engine	7,439,433	47.1
Library	2,277,215	14.4
Other	2,149,130	13.6
Not specified	1,078,661	6.8
WC other	946,696	6.0
WC home	829,546	5.3
Citation service	578,133	3.7
GoodReads.com	250,293	1.6
Wikipedia	155,427	1.0
OCLC services	95,193	0.6

TABLE 6. Distribution of referrers for top 10 countries (percentage of sessions originating from each country that come from each referrer).

	Search engine	Library	Other	Not specified	WC other	WC home	Citation service	GoodRead	Wikipedia	OCLC service
US	29.8	27.5	13.4	8.0	4.1	7.5	6.9	1.4	0.8	0.6
China	50.3	13.9	13.2	1.7	18.8	1.3	0.1	0.1	0.3	0.3
Canada	30.2	4.5	40.1	8.4	3.2	5.7	5.2	1.4	0.8	0.4
UK	58.7	5.5	10.1	9.0	5.9	5.5	1.1	1.7	2.0	0.5
Germany	59.6	0.7	14.5	12.7	6.2	3.9	0.3	0.4	1.1	0.6
France	67.0	0.9	12.2	8.3	5.4	3.2	0.2	0.3	1.7	0.8
India	71.1	1.7	8.0	3.4	3.8	1.8	0.3	6.7	2.1	1.2
Italy	77.5	1.1	8.1	5.1	3.6	2.5	0.3	0.3	0.8	0.7
Indonesia	69.9	0.3	10.8	2.3	10.8	0.7	0.1	4.3	0.4	0.4
Spain	71.2	2	10.6	6.4	4.8	2.6	0.3	0.8	1.1	0.7

instruction sessions for demonstration purposes. This last work-task can be distinguished from the previous three in that it incorporates no subsidiary search-task.

Several work-tasks were described by students and academics. All of the academics and several postgraduate students spoke generally of using the system to aid their research. The responses of undergraduate students to the question of why they accessed the system indicated that it was almost without exception for the purposes of aiding a defined academic assignment such as an essay or presentation. Although it was clear that most viewed WorldCat.org as primarily an academic or professional resource, a small number of participants from all groups also mentioned using the system for leisure purposes, either as a means of finding books to read for pleasure, or in support of their own hobbies. It is an acknowledged limitation of this study that the primarily qualitative data were gathered from academic and professional users of the system. We suggest therefore that a wider range of leisure-related work-tasks would be revealed through a more in-depth study of recreational users.

Search-tasks. Results from all three phases of the project revealed three distinct classes of search-task: searches for a known-item (e.g., to determine the closest library holding a particular title), searches for an unknown-item (e.g., checking for new publications on a particular topic), and searches for institutional information (e.g., to find the address of a library).

Focus group interview participants described a wide range of known-item search tasks. Among the most commonly mentioned, particularly by librarians and booksellers, was the task of determining the bibliographic details of an item. A number of variations of this type of task were described. Participants told of using the system to check bibliographic details as part of a standard validation process (“We use WorldCat.org to verify if the bibliographic details are correct,” NZ public librarian), or confirming details about which the searcher had some doubt. A number of librarians also spoke of using the system to confirm a reference based on incomplete or incorrect information. Interestingly, although a number of academic librarians described occasions when they had used WorldCat.org to verify a reference given to them by a patron, no students mentioned using the system for this purpose.

Another very frequently mentioned known-item search-task was related to determining locations where a particular item is held. Students, librarians, and academics all described situations in which they used the “Find a Copy in the Library” function from WorldCat.org to ascertain which library or libraries held the item (“It’s a tool for locating things,” UK historian, “WorldCat.org is often the best option for locating a book outside the library,” US academic Librarian). Some participants described using this service as a means of determining libraries to which they could submit ILL requests. This particular search task was one that could be identified clearly in the transaction logs because it was possible to identify instances of a user clicking on a link to a library site from the list presented by the “Find a copy in the library” feature. Overall, 5.81% of sessions were found to include at least one such click ($n = 918,698$), which equates to almost half a million such sessions per month. Further analysis, however, reveals significant variations in the proportion of sessions from different referrer types that include this activity (Table 7). We note that although almost a quarter of sessions referred to WorldCat.org by a library include such an action, only a tiny proportion (0.05%) of search engine referrer sessions do so. Similarly, there is significant geographic variation, with only the United States (10.1%) and Canada (8.2%) having greater than 1% of sessions including the action.

TABLE 7. Sessions including a click on a link to a library holding the item by referrer type.

Referrer type	Sessions including click on link to a library holding the item	
	% of sessions from referrer	Total number of sessions
Library	24.54	558,925
Other	13.83	297,275
Citation service	7.94	45,885
OCLC services	1.24	1,185
Not specified	0.77	8,288
WC home	0.32	2,659
WC other	0.09	855
Search engine	0.05	3,481
Goodreads.com	0.04	101
Wikipedia	0.03	44
All referrers	5.81	918,698

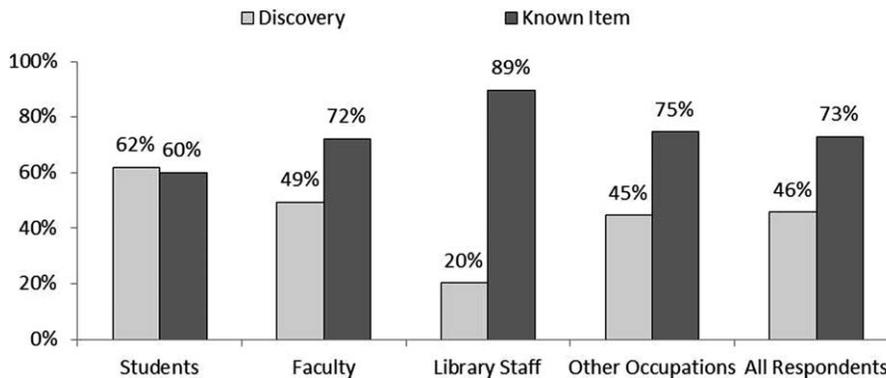


FIG. 4. Respondents engaged in known-item or discovery search tasks.

Another important use of WorldCat.org described by librarians and booksellers was using the system to determine the number of libraries holding a particular item. For librarians, this often was spoken of as aiding decisions relating to acquisitions. Some librarians spoke generally about comparing their own collections to those of other libraries: “Collection overlap is a key focus area” (Australian academic librarian). There was a strong sense here that knowing whether other local libraries held an item would influence the likelihood of acquisition.

Other participants were seeking a single specific edition of a work: “I was looking for a specific edition of *Moby Dick* that I’d read about and knew had interesting illustrations. I was able to find it on WorldCat” (US graduate student). Academics and students were particularly interested in locating electronic versions of a particular book, something made clear not only by their own comments (“I’m checking WorldCat.org to check if there’s a digital version,” UK historian; “Quite often I go to WorldCat.org to see if there’s an ebook that I can try and get access to,” US undergraduate student), but also from the comments of librarians who had assisted them:

Students are very interested in the format. They almost always want instant access, and feel electronic versions can provide that. If a student comes up to me at the desk and asks about an item that we don’t have in electronic form, WorldCat.org is somewhere I can go to see what e-versions are out there. (NZ academic librarian)

Finding unknown-items also emerged as an important use of the system. As one UK undergraduate student put it: “I think that’s my primary use of WorldCat.org — to find things I did not know existed.” Analysis of the data generated from the focus groups revealed a range of unknown-item search tasks undertaken by participants on WorldCat.org. It is instructive to note here that the range of search-tasks classed as unknown-item go beyond what reasonably might be considered topical-searches. A good example of this relates to the identification of unknown titles by a known author. This was spoken of by librarians, historians, and students as an effective and commonly used means of discovering useful resources.

Topic searches nonetheless represented the most frequently mentioned form of unknown-item search. The typical approach to these searches was summed up by one student: “I put in keywords and find useful things” (UK graduate student). Students and librarians frequently described situations where they used WorldCat.org to identify multiple items on a topic:

I mostly use [WorldCat.org] to try to find initial sources of material for an assignment. I had to find sources about rescue helicopters and there were quite a few books about them on WorldCat.org. (US graduate student)

Academic librarians also spoke of directing students seeking additional material on a topic to WorldCat.org: “we often suggest WorldCat.org to students after they’ve used our own catalog, particularly for topic searches” (US academic librarian). It was also apparent that for some participants, WorldCat.org was perceived as particularly useful for more obscure subject areas.

Sometimes participants described search tasks that did not require the identification of multiple resources, but just one unknown-item. In these cases the searcher was most often looking for a single item on a topic that met some strict criteria relating to audience level or specific subject:

A Professor wanted to read a story to his son’s 2nd grade class. He wanted a book on kayaking suitable for 7 year olds. To maintain street cred I checked WorldCat.org and was able to find something appropriate. (US academic librarian)

Students described in general terms how they sometimes found it useful to try and find items that were similar to resources that had previously proved useful, and more specifically spoke of occasions when they had been required to find alternatives to a known item, for example, when the item they sought was on loan. Descriptions of topic searches also related to finding everything available on a given topic. Academic librarians spoke of how PhD students and academics viewed WorldCat.org as an ideal system for ensuring the completeness of their searches. For PhD students this

TABLE 8. Sample session task-type coding by referrer type.

Task-type	WC home	Library	Search engine	Combined
Known-item	69.2%	62.4%	59.0%	63.6%
Unknown-item	11.9%	20.8%	24.0%	18.8%
Known-item and unknown-item	2.2%	0.0%	0.0%	0.7%
Author	2.9%	2.0%	3.0%	2.6%
WC account	0.2%	0.0%	0.0%	0.1%
Library info	0.2%	0.0%	0.0%	0.1%
Not classified	13.3%	14.9%	14.0%	14.1%

was often to make sure they had identified all the literature in their area, whereas for academics it was frequently related to ensuring nobody had covered the precise subject of their research.

Survey respondents also were asked about their reasons for accessing the system, with a general distinction made between the goals of locating a specific known-item in the catalog, and broader topic searches. Figure 4 shows the results for this question (note that respondents were able to select more than one goal if their session encompassed both types of task). Library staff were found to be much more likely to be undertaking some form of known-item search, with 89.5% ($n = 656$) respondents from this group engaged in this activity, compared with 60.4% ($n = 634$) of students.

The proportion of respondents engaged solely in known-item tasks is even more revealing, in that over three-quarters (77.1%, $n = 565$) of library staff responding to the survey were determining either the location or some bibliographic information about a known-item. In contrast, fewer than half of students said they were only conducting a known item search (37.1%, $n = 389$). These results were statistically significant, $\chi^2(3, N = 2,918) = 279.80, p < .001$, with a large effect size (Cramer's $V = 0.310$).

Analysis of the sample session from the transaction log files also attempted to quantify the proportion of users engaged in different types of search task. Table 8 presents the distribution of task type for each of the three referrer types. In total, 169 sessions (14.1%) proved impossible to confidently code. The majority of sessions for each referrer type were coded as known-item, with 63.6% ($n = 763$) of the combined sample set assigned this code. Results were relatively consistent for each referrer type, with no statistically significant differences. Unknown-item tasks represented the next largest proportion of sessions, with almost a fifth (18.8%, $n = 226$) of all sample sessions allocated this code. Differences were observed in the number of unknown-item sessions for each referrer, with almost a quarter of search engine sessions (24%, $n = 96$) ascribed the code compared to 11.9% of WorldCat.org homepage sessions ($n = 47$) and 20.8% of library sessions ($n = 83$). These results were found to be statistically significant, $\chi^2(2, N = 246) = 3.28, p < .001$. All other codes were very rarely assigned, with author searches representing fewer than 3% of all sessions ($n = 31$), and the other codes combined accounting for fewer than 1% ($n = 11$).

A number of participants told of occasions when they had used WorldCat.org to ascertain information about libraries. Several librarians spoke of using WorldCat.org to find the address of a library, usually for the purpose of correspondence. Students also spoke of using the system to find the address of a library, typically in order to facilitate a visit. Librarians also described using the system to determine other libraries' ILL policies. Several participants spoke of undertaking more sophisticated search-tasks on the system that were related to understanding individual library specializations. Librarians tended to use such searches as way of staying up to date with collection development policies at rival institutions, and to gather information that might influence future collection development decisions. The only academic to mention this type of task explained that they were keen to understand which libraries would be most beneficial to visit.

Discussion

This paper has explored the users and uses of WorldCat.org using a three-phase mixed-methods approach. Three research questions were posed, which we discuss now.

The first research question related to the demographics of WorldCat.org users. The age and gender of users were found to match closely the results of the 2009 WorldCat.org study (Calhoun et al., 2009), although it must be noted that the survey respondents are not necessarily representative of the wider WorldCat.org user base. Both the transaction log analysis and survey also revealed the wide geographic spread of WorldCat.org users. Although focus group interview participants from the UK and Australasia commented that the system could at times feel US-centric, a consequence no doubt of its origins as a North American service, our results demonstrate that it can with some justification now be termed a global service, with almost half of all traffic originating outside the United States and Canada. Because numerous studies have shown that cultural factors affect interactions with systems, including general search behavior (Zoe & DiMartino, 2000), query reformulation (Jesper, Clough, & Hall, 2013), and information-seeking behavior (Ford, Miller, & Moss, 2001), we suggest that significant attention should be paid to ensuring that the system best meets the needs of users from around the world.

The survey results also indicate three primary user groups—librarians, students, and academics—which serves to validate the selection of focus group interview participants. These again match the key user groups found in the small amount of literature available on WorldCat.org users, and union catalogs in general (e.g., Goodale & Clough, 2012; Hartley & Booth, 2006). Compared directly with the results of the 2008 survey (Calhoun et al., 2009), we note a greater proportion of student respondents to our survey (2008 = 16%, 2012 = 36%), and a smaller proportion of librarians (2008 = 36%, 2012 = 23%). The results of the focus groups suggest that this increase may in part be due to increased awareness of the service for student groups,

TABLE 9. Emerging taxonomy of WorldCat.org work and search-tasks.

Work tasks		
Level 1	Level 2	
Academic	Essay/Assignment	
	Research	
Leisure	Hobbies	
	Reading for Pleasure	
Professional	Acquisitions/Collection	
	Development	
	Cataloging	
	Inter-Library Loan	
	Instruction/Training	
	Reading-List Development	
	Valuation	
Search tasks		
Level 1	Level 2	Level 3
Institutional information	Location	
	Policies	
	Specializations	
Known-item	Bibliographic Details	
	Editions	
	Format	
	Location	
	Holdings	
	Citation	
Unknown-item	Related	Author
		Manifestation
		Similar item
	Topic	Completeness
		Monitoring
		Multiple items
		Single item

particularly through the use of links to WorldCat.org from institutional catalogs.

The second research question addressed the issue of how users were being referred to WorldCat.org. The analysis conducted on the full WorldCat.org logs included the assignment of a referrer type to each session in the log, with results showing that almost half of all sessions originated from a search engine results page, and a further 14% coming from library pages. The log analysis also revealed differences in behavior and levels of system interaction between sessions originating from different referrer types, most significantly in the way that users who started directly at the homepage generally spent longer on the system, and were much more likely to execute queries.

Perhaps the most striking finding from the transaction log analysis was the large number of sessions originating from search engine referrals that consisted of no further engagement with the system after arriving at the site. The nature of the data makes it impossible to accurately determine what activity these sessions represent. Such sessions only can be said to represent a user executing a query on a search engine, and clicking on a WorldCat.org link from the search engine results page. This link takes them directly to an item record page. Depending on the nature of their search task, it is feasible that viewing this single page satisfies their information

need (e.g., if they are seeking some bibliographic data about an item). Alternatively, it is possible that such users are undertaking tasks for which WorldCat.org is unsuited (e.g., purchasing a book, seeking reviews). While the overall proportion of sessions including a click on a link to a library holding the item was found to be 5.8%, exactly in line with Nilges previous estimate of 4% to 6% (2006), it is notable that only a tiny proportion (<1%) of search engine referred sessions included such an action. Although we cannot be certain of the reasons for the low click-through rate, one possible reason was suggested in a number of focus group interview sessions, namely, accessing full-text online. We suggest that a significant proportion of users referred to WorldCat.org from search engines are likely to be seeking full-text online versions of the object of their search. A number of studies have reported that web users expect instant and unimpeded access to such material (Ballard & Blaine, 2011; Markey, 2007; Neal, 2009), and they are perhaps unlikely to view links to local library catalogs as a productive means of facilitating this access.

Thus, these results can be said to offer limited support to the notion suggested by Dempsey (2006) that union catalogs offer an effective means of exposing individual library holdings. On one hand, we note that search engine referrals drive a high volume of traffic to WorldCat.org, particularly from outside the United States. However, we also note that these referrals very rarely result in a click-through to an individual library. This is in stark contrast with referrals from other library services, almost one in four of which result in such a click-through. Our results suggest therefore that the greatest success in exposing collections has been found through links to WorldCat.org from individual catalogs. These facilitate the sort of searching described frequently in the focus group interview sessions, whereby users seeking a specific item not available from their own library are able to use WorldCat.org to identify copies held in other libraries, and subsequently request through ILL or collect in person. Thus, the system does successfully facilitate access “above the level of individual libraries” (Dempsey, 2006), but only usually for users already engaging with a library system.

Our third research question examined the purposes for which users accessed WorldCat.org. In developing taxonomies of work and search tasks, it must be acknowledged that other populations with potentially relevant input were not investigated. Several participants described their use of the system for leisure purposes, allowing for the generation of a category of Leisure-related work-tasks. Participants also included rare book sellers, who were able to describe their professional reasons for using the site, but it is clear that their needs are highly specialized, and unlikely to represent use cases for a host of other professions identified as users by the Phase 2 survey. Thus, the emergent work- and search-task taxonomies presented in Table 9 are potentially incomplete; while they represent a robust representation of student and librarian needs, and therefore capture the most common use cases, there is potential for expansion to encompass uses by other professions and leisure users.

There is very little literature against which to benchmark these findings. Although Goodale and Clough's four use-scenarios of the SEARCH25 catalog (2012) are all represented by this taxonomy, Slone's notion of an Area search (2000) is not included because it is only applicable in circumstances when the user is searching a catalog with the intention of determining the location of an item within the physical library. In general, the taxonomy provides a more detailed breakdown of the "Known-item" and "Discovery" purposes identified by Calhoun et al. (2009).

The majority of search-tasks undertaken on WorldCat.org are certainly for known-items. However, the coding of the sample log sessions offers some mechanism for estimating the number of sessions involving unknown-item search tasks in the wider logs: 18.8% of sessions including a query were found to include an unknown-item search, representing around 3% of all sessions. This figure is significantly lower than those found in prior studies of both union and institutional catalogs (e.g., Goodale & Clough, 2012; Larson, 1991; Slone, 2000). Some explanation for this can be found in the results of our focus group interviews. Several participants described looking for resources on a topic first using their institutional catalog, then a local or national union catalog, before accessing WorldCat.org. As one historian put it: "I'd purposely use WorldCat if I'd exhausted other major resources." It is reasonable to imagine that a large number of unknown-item search-tasks are resolved at the institutional or local level, resulting in a lower number of such queries being executed in WorldCat.org. It is important to note that although the proportion of unknown-item searches may be low, the high volume of traffic coming to the site means that unknown-item searching occurs in around 250,000 sessions each month. Thus, although supporting unknown-item search may not be WorldCat.org's primary goal, there appears to be a significant number of users who do use the system for this purpose, and thus motivation for OCLC to explore potential means of improving the discovery process.

Conclusions

The changing nature of digital library services and the needs and expectations of users requires that service providers continue to assess and update their services and systems. In this paper we have carried out an in-depth study of the users and uses of WorldCat.org, the world's largest bibliographic database and global union catalog using a mixed-methods approach consisting of focus group interviews, a pop-up survey, and transaction log analysis. It is clear from the findings that WorldCat.org is used by a large and diverse user population. Although the two largest single groups of users are librarians and students, with academics also constituting a significant proportion of the whole, survey respondents included professions as diverse as gardeners, actors, and accountants. Analysis of the log files also revealed the diversity of geographic locations from which users access the site. Although the majority of traffic originates from North

America, millions of sessions were found to originate from countries on all continents. Thus, although the typical user might be a US librarian or student, it is clear that WorldCat.org must cater to a vast range of cultural and linguistic needs. Our findings also show that search engine referrals account for almost half of all traffic arriving at WorldCat.org, but that these sessions typically comprise very little further interaction with the system. In the future we hope to investigate these sessions in order to better understand whether they represent easily resolved search-tasks, mistaken clicks, or some other use case.

We also present an emerging taxonomy of WorldCat.org work and search tasks, based on analysis of focus group interviews with 120 users on three continents. We acknowledge that although this taxonomy provides a robust representation of the motivations of librarian, student, and academic users, it as fully reflects the needs of other users of WorldCat.org. Further investigations of nonacademic users, and users of other union catalogs, would serve to validate and expand the taxonomy. Our results do support the notion that union catalogs are primarily used for known-item searches, while noting that the sheer volume of traffic arriving at WorldCat.org means that the relatively small proportion of unknown-item searches still represent a large number of sessions. Better understanding users' reasons for accessing a system allows for a more robust evaluation of how well that system performs, and further work investigating the extent to which the features and functionality of WorldCat.org support users in their information seeking is already underway.

Finally, our analysis of clicks on links to individual libraries holding an item suggests that although WorldCat.org is highly successful at connecting users referred from institutional library catalogs to other libraries holding a sought item, users arriving from a search engine referral are much less likely to connect to an individual library. Integrating institutional library collections with popular web-scale discovery tools, particularly search engines, therefore remains an ongoing and important challenge.

References

- Alam, N., & Pandey, P. (2012). GeoTheses: Development of union catalog of Indian geoscience theses using GSDL. *The Electronic Library*, 30(4), 456–468.
- Allen, B. (1996). *Information tasks: Toward a user-centered approach to information systems*. Orlando, FL: Academic Press.
- Ballard, T., & Blaine, A. (2011). User search-limiting behavior in online catalogs: Comparing classic catalog use to search behavior in next-generation catalogs. *New Library World* 112(5/6), 261–273.
- Bazeley, P., & Kemp, L. (2012). Mosaics, triangles, and DNA: Metaphors for integrated analysis in mixed methods research. *Journal of Mixed Methods Research*, 6(1), 55–72.
- Berger, W.K., & Hines, R.W. (1984). What does the user really want? The library user survey project at Duke University. *Journal of Academic Librarianship*, 20(5-6), 306–309.
- Bernstein, J.H. (2006). From the ubiquitous to the non-existent: a Demographic Study of OCLC WorldCat. *Library Resources & Technical Services*, 50(2), 79–90.

- Bertot, J.C., Berube, K., Devereaux, P., Dhakal, K., Powers, S., & Ray, J. (2012). Assessing the usability of WorldCat local: Findings and considerations. *The Library Quarterly*, 82(2), 207–221.
- Blecic, D.D., Bangalore, N.S., Dorsch, J.L., Henderson, C.L., Koenig, M.H., & Weller, A.C. (1998). Using transaction log analysis to improve OPAC retrieval results. *College & Research Libraries*, 59(1), 39–50.
- Boston, T., Rajapatirana, B., & Missingham, R. (2009). Libraries Australia: Simplifying the search experience. National Library of Australia Staff Papers. Retrieved from <http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1231>.
- Breeding, M. (2010). The state of the art in library discovery. *Computers in Libraries*, 30(1), 31–34.
- Broder, A. (2002). A taxonomy of web search. *SIGIR Forum*, 36(2), 3–10.
- Burnhill, P., & Law, D. (2005). SUNCAT rising: UK serials union catalogue to assist document access. *Interlending & Document Supply*, 33(4), 203–207.
- Bystrom, K., & Hansen, P. (2005). Conceptual frameworks for tasks in information studies. *Journal of the American Society for Information Science and Technology*, 56(10), 1050–1061.
- Calhoun, K.S., Cantrell, J., Gallagher, P., & Cellantani, D. (2009). Online catalogs: What users and librarians want. Dublin OH: OCLC.
- Chen, X. (2012). Google Books and WorldCat: A comparison of their content. *Online Information Review*, 36(4), 507–516.
- Connaway, L.S. (2007). Mountains, valleys, and pathways: Serials users' needs and steps to meet them. Part I: Identifying serials users' needs: Preliminary analysis of focus group and semi-structured interviews at colleges and universities. *Serials Librarian*, 52(1/2), 223–236.
- Connaway, L.S., & Powell, R.R. (2010). Basic research methods for librarians. Santa Barbara, CA: ABC-CLIO.
- Connaway, L.S., O'Neill, E.T., & Prabha, C. (2006). Last copies: What's at risk?. *College & Research Libraries*, 67(4), 370–379.
- Connaway, L.S., Wilcox, D.J., & Searing, S.E. (1997). Online catalogs from the users' perspective: The use of focus group interviews. *College & Research Libraries*, 58(5), 403–420.
- Cousins, S. (1999). Virtual OPACs versus union database: Two models of union catalog provision. *The Electronic Library*, 17(2), 97–103.
- Craven, J., Johnson, F., & Butters, G. (2010). The usability and functionality of an online catalogue. *Aslib Proceedings*, 62(1), 70–84.
- Dempsey, L. (2006). Libraries and the long tail: Some thoughts about libraries in a network age. *D-Lib Magazine* 12(4). Retrieved from <http://www.dlib.org/dlib/april06/dempsey/04dempsey.html>.
- Feather, J., & Sturges, P. (Eds.). (2003). *International encyclopedia of information and library science*. West Sussex, UK: Routledge.
- Ford, N., Miller, D., & Moss, N. (2001). The role of individual differences in Internet searching: An empirical study. *Journal of the American Society for Information Science and Technology*, 52(12), 1049–1066.
- Gisbergen, M.S.V., Most, J.V.D., & Aelen, P. (2007). Visual attention to online search engine results. Market Research Agency De Vos & Jansen.
- Goodale, P., & Clough, P. (2012). Report on Evaluation of the SEARCH25 Demo System. University of Sheffield: Sheffield, UK.
- Gorman, M. (2007). Union catalogs: Their role in library networking and their continued relevance in a digital age. In Conference: Libraries: Networking for National Development, Libraries, Jamaica, November (pp. 22–23).
- Gradmann, S. (2004). The Cathedral and the Bazaar, Revisited: Union Catalogs and Federated WWW Information Services. In A. Lass & R.E. Quandt (Eds.), *Union catalogs at the crossroad* (pp. 67–87). Hamburg, Germany: Hamburg University Press.
- Hartley, R.J., & Booth, H. (2006). Users and union catalogs. *Journal of Librarianship and Information Science*, 38(1), 7–20.
- Hert, C.A. (1996). User goals on an online public access catalog. *Journal of the American Society for Information Science*, 47(7), 504–518.
- Hider, P. (2004). The bibliographic advantages of a centralised union catalog for ILL and resource sharing. *Interlending & Document Supply*, 32(1), 17–29.
- Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, 28(3), 407–432.
- Jansen, B.J., Booth, D.L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing and Management*, 44, 1251–1266. doi:10.1016/j.ipm.2007.07.015
- Jesper, S., Clough, P., & Hall, M. (2013). Regional effects on query reformulation patterns. Paper presented at The International Conference on Theory and Practice of Digital Libraries (TPDL 2013) (pp. 382–385). Valletta, Malta.
- Johnson, R.B., & Onwuegbuzie, A.J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Jones, R., & Klinkner, K.L. (2008). Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs. Paper presented at CIKM'08 (pp. 699–708). Napa Valley, CA: ACM Press.
- Kan, M.-Y., & Poo, D.C.C. (2005). Detecting and Supporting Known Item Queries in Online Public Access Catalogs. Paper presented at ACM/IEEE Joint Conference on Digital Libraries, JCDL. Denver, CO: ACM Press.
- Kilgour, F.G. (2001). Known-item online searches employed by scholars using surname plus first, or last, or first and last title words. *Journal of the American Society for Information Science and Technology*, 52(14), 1203–1209.
- Kitalong, K.S., Hoepfner, A., & Scharf, M. (2008). Making sense of an academic library web site: Toward a more usable interface for university researchers. *Journal of Web Librarianship*, 2(2-3), 177–204.
- Larsen, K. (2007). Bibliotek.dk: Opening the Danish union catalog to the public. *Interlending & Document Supply*, 35(4), 205–210.
- Larson, R. (1991). The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science and Technology*, 42(3), 197–215.
- Lass, A., & Quandt, R.E. (Eds.). (2004). *Union catalogs at the crossroad*. Hamburg, Germany: Hamburg University Press.
- Lavoie, B., Connaway, L.S., & Dempsey, L. (2005). Anatomy of aggregate collections: The example of Google Print for libraries. *D-Lib Magazine*, 11(9). Retrieved from <http://www.dlib.org/dlib/september05/lavoie/09lavoie.html>
- Lee, J.H., Renear, A., & Smith, L.C. (2006). *Known-item search: Variations on a concept*. Paper presented at the 69th ASIS&T Annual Meeting, 43, 1–17. Austin, TX.
- Lewandowski, D. (2010). Using search engine technology to improve library catalogs. *Advances in Librarianship*, 32, 35–54.
- Little, G. (2012). Where are you going, where have you been? The evolution of the academic library web site. *The Journal of Academic Librarianship*, 38(2), 123–125.
- Markey, K. (2007). The online library catalog: Paradise lost and paradise regained? *D-Lib Magazine*, 13(1), 4.
- Miksa, F. (2009). Information organization and the mysterious information user. *Libraries & The Cultural Record*, 44(3), 343–370.
- Mittal, R. (2011). The National Union Catalogue of Scientific Serials in India (NUCSSI) on the web. *Interlending & Document Supply*, 39(1), 53–60.
- Neal, J.G. (2009). What do users want? What do users need? W (h)ither the academic research library? *Journal of Library Administration*, 49(5), 463–468.
- Nilges, C. (2006). The online computer library center's open WorldCat program. *Library trends*, 54(3), 430–447.
- Ockuly, J. (2003). What clicks? An interim report on audience research. Paper presented at the Museums and the Web Annual Conference. Charlotte, NC: Archives & Museum Informatics.
- OCLC. (2015). A global library resource [Website]. Retrieved from: <http://www.oclc.org/worldcat/catalog.en.html>.
- Perrault, A.H. (2002). *Global collective resources: A study of monographic bibliographic records in WorldCat*. Tampa, FL: University of South Florida Scholar Commons.

- Peters, T.A. (1993). The history and development of transaction log analysis. *Library Hi Tech* 11(2), 41–66.
- Phippen, A., Sheppard, L., & Furnell, S. (2004). A practical evaluation of Web analytics. *Internet Research: Electronic Networking Applications and Policy*, 14, 284–293.
- Rose, D.E., & Levinson, D. (2004). Understanding user goals in web search. Paper presented at the 13th International World Wide Web Conference (WWW 2004) (pp. 13–19). New York: ACM Press.
- Shneiderman, B., Byrd, D., & Croft, W.B. (1997). Clarifying search: A user-interface framework for text searches. *D-lib Magazine*, 3(1), 18–20.
- Slone, D.J. (2000). Encounters with the OPAC: On-Line searching in public libraries. *Journal of the American Society for Information Science*, 51(8), 757–773.
- Teets, M., & Goldner, M. (2013). Libraries' role in curating and exposing big data. *Future Internet*, 5(3), 429–438.
- Toms, E.G. (2011). Task-based information searching and retrieval. In I. Ruthven & D. Kelly (Eds.), *Interactive information seeking, behavior and retrieval* (pp. 43–59). London, UK: Facet Publishing.
- Vakkari, P. (2003). Task-based information searching. *Annual Review of Information Science and Technology*, 37(1), 413–464.
- Von Seggern, M., & Young, N. J. (2003). The focus group method in libraries: Issues relating to process and data analysis. *Reference Services Review*, 31(3), 272–284.
- Wakeling, S. (2015). Establishing user requirements for a recommender system in an online union catalogue: An investigation of WorldCat.org (Doctoral dissertation, University of Sheffield, UK).
- Wakeling, S., & Clough, P. (2015). Integrating mixed-methods for evaluating information access systems. Paper presented at CLEF'15 (pp. 306–311). Cham, Switzerland: Springer.
- Wilkinson, S. (2004). Focus group research. In D. Silverman (Ed.), *Qualitative research: Theory, method, and practice* (pp. 177–199). Thousand Oaks, CA: Sage.
- Wilson, T. (1999). Models in information behavior research. *Journal of Documentation*, 55(3), 249–270.
- Yardley, L. (2008). Demonstrating validity in qualitative psychology. *Qualitative Psychology: A Practical Guide to Research Methods*, 2, 235–251.
- Yee, M.M., & Layne S.S. (1998). *Improving online public access catalogs*. Chicago, IL: American Library Association.
- Zhang, Y., & Wildemuth, B.M. (2009). Qualitative analysis of content. In B. Wildemuth (Ed.), *Applications of social research methods to questions in information and library science* (pp. 308–319). Westport, CT: Libraries Unlimited.
- Zoe, L.R., & DiMartino, D. (2000). Cultural diversity and end-user searching: An analysis by gender and language background. *Research Strategies*, 17(4), 291–305.