University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, UNL Libraries | Libraries at University of Nebraska-Lincoln

2-16-2016

OAI-PMH Harvested Collections & User Engagement

DeeAnn Allison
University of Nebraska-Lincoln, dallison1@unl.edu

Follow this and additional works at: https://digitalcommons.unl.edu/libraryscience

Part of the Cataloging and Metadata Commons

Allison, DeeAnn, "OAI-PMH Harvested Collections & User Engagement" (2016). *Faculty Publications, UNL Libraries*. 337.
https://digitalcommons.unl.edu/libraryscience/337

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Faculty Publications, UNL Libraries | Libraries at University of Nebraska-Lincoln

2-16-2016

OAI-PMH Harvested Collections & User Engagement

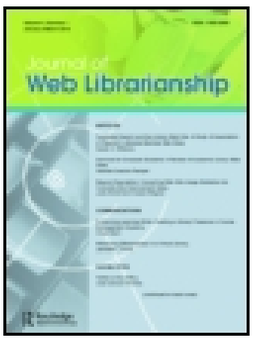
DeeAnn Allison
University of Nebraska-Lincoln, dallison1@unl.edu

Follow this and additional works at: http://digitalcommons.unl.edu/libraryscience

 Part of the Cataloging and Metadata Commons

Allison, DeeAnn, "OAI-PMH Harvested Collections & User Engagement" (2016). *Faculty Publications, UNL Libraries.* Paper 337.
http://digitalcommons.unl.edu/libraryscience/337

This Article is brought to you for free and open access by the Libraries at University of Nebraska-Lincoln at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, UNL Libraries by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.



Routledge
Taylor & Francis Group

Journal of Web Librarianship

ISSN: 1932-2909 (Print) 1932-2917 (Online) Journal homepage: http://www.tandfonline.com/loi/wjwl20

OAI-PMH Harvested Collections and User Engagement

DeeAnn Allison

To cite this article: DeeAnn Allison (2016): OAI-PMH Harvested Collections and User Engagement, Journal of Web Librarianship, DOI: 10.1080/19322909.2015.1128867

To link to this article: http://dx.doi.org/10.1080/19322909.2015.1128867

Published online: 16 Feb 2016.

Submit your article to this journal

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=wjwl20

Download by: [University of Nebraska, Lincoln] **Date:** 17 February 2016, At: 04:44

JOURNAL OF WEB LIBRARIANSHIP
http://dx.doi.org/10.1080/19322909.2015.1128867

Routledge
Taylor & Francis Group

ARTICLE

OAI-PMH Harvested Collections and User Engagement

DeeAnn Allison

University of Nebraska-Lincoln Libraries, Lincoln, Nebraska, USA

ABSTRACT

Discovery tools are used in libraries to bring together books, articles, and other resources. Research has focused on user and librarian evaluation of these tools, but there are few evaluations of non-book and non-article sources. Discovery tools can also include metadata for local collections harvested through the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Creating these harvests can be time consuming for staff, so it is important for libraries to understand if and how patrons use these records. The University of Nebraska-Lincoln Libraries (UNL Libraries) harvests metadata from local collections into the Encore discovery tool. A study was conducted to analyze patron use of OAI-harvested records. This study analyzed usage data for harvested collections obtained from different discovery sources and referrals through Encore. Google Analytics was used to evaluate searcher behavior differences between content referred through Encore and other referrals. Although discovery through Encore did not result in high numbers of traffic, there is evidence that patrons who discover records through Encore take more time looking through records than patrons using other discovery methods. This increase in time is a measure of engagement and may be reason enough for libraries to consider adding OAI-harvested collections to their discovery tool.

ARTICLE HISTORY
Received 9 October 2015
Accepted 2 December 2015

KEYWORDS
CONTENTdm; discovery; Encore; engagement; Google Analytics; OAI-PMH

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

Introduction

Unmediated searching for research information has become normal in academic institutions. However, scholars are often frustrated with isolated databases that require them to repeat their searches using different query methods. The discovery tool is an attempt to answer this criticism by integrating resources with traditional catalog entries to provide a single point for searching. This integration can include more than just articles with the addition of locally developed resources that are frequently hidden in repositories and disconnected from the catalog. There has been much discussion and evaluation of the integration of articles and monographs into discovery tools, but not very much has been written about Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) harvested content. This article

CONTACT DeeAnn Allison dallison1@unl.edu 317 Love Library, University of Nebraska-Lincoln, Lincoln, NE 68588-0410.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/wjwl

© 2016 DeeAnn Allison

reports on a study of differences between the way users engage with OAI-PMH harvested content discovered through Encore compared with content accessed from other sources.

Literature review

A 2015 paper prepared for the National Information Standards Organization (NISO) by Marshall Breeding (2015, 49) discussed the current challenges and future of resource discovery. The paper highlighted the need for transparency in content coverage and forecasted continued dominance of the commercial sector in influencing the content and functionality of discovery tools. Although there is general agreement about required fields for searching, there is a difference between the indexing provided by commercial services for article inclusion and other content such as images and e-books. Intermixing content poses challenges for discovery tools in displaying and ranking results in an intuitive manner that patrons can understand.

There is no doubt that vendors play a major role in identifying the content and presentation of commercial discovery products. Librarians are beginning to evaluate the usefulness of these tools to identify strengths and weaknesses. In a 2012 survey of the literature on discovery tools, Thomsett-Scott and Reese (2012, 123–43) provided an excellent summary of articles on discovery tools. Librarians offered differing opinions about the strengths and weaknesses of different tools. The strengths of the tools center on the single search option, facets for drilling down, and added features including citation management, reviews, and rating information. The weak areas include too many results, problems with relevancy matching, missing content, and the ongoing need for user instruction. In a 2013 study by Stephen Bull, Edward Craft, and Andrew Dodds, patrons were asked to compare the catalog against a new Primo discovery tool. Their findings revealed that 55 percent preferred the discovery tool over the catalog with two-thirds of the undergraduates preferring the tool (2014, 137–66).

A usability study by Mireille Djenno and colleagues of researcher preferences between WorldCat Local and Summon highlighted the need for librarians to rethink how discovery tools are evaluated. They noted, “In some instances, the tasks researchers asked participants to complete were not necessarily tasks that participants would be inclined to perform in a real-world situation” (2014, 277). This observation points to a need for researchers to look at discovery as a part of the research process. As Condit Fagan and her collaborators (2012, 104) discussed, “although discovery tools challenge libraries to think not only about access but also about the best research pathways for users, they provide users with a search that more closely matches their expectations.” Andrew Asher, Lynda Duke, and Suzanne Wilson agreed, emphasizing that discovery settings that enhance the way students are guided were of paramount importance in selecting a tool (2013, 464–88). Clearly, librarians must identify ways to capture and evaluate user interactions as they evaluate the effectiveness of discovery tools.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

There is evidence that the implementation of discovery tools results in higher use of library resources. In one comparison study of EBSCO Discovery Service (EDS), ExLibris' Primo, Serial Solutions' Summon, and OCLC's WorldCat, 33 libraries from around the world were evaluated for publisher-hosted journal usage (Levine-Clark, McDonald, and Price 2014, 249–56). The authors found that discovery service implementation was a strong predictor of increased journal use. Kristin Calvert reported on the results of a study on usage patterns in EBSCO's EDS service and concluded, "EBSCO Discovery Service has undoubtedly changed user behavior to better connect the patron to the library's e-journals and to increase use of full text and abstracts in EBSCOhost databases" (2015, 96).

Since use of research content increases through the use of a discovery tool, the following question arises: How does use of a discovery tool impact the use of other content? In addition to books and electronic resources, it is possible to integrate other resources into a discovery tool. Susan Johns-Smith (2012, 17–23) described the workflow for creating an OAI-enriched discovery tool and emphasized the importance of identifying content to include. It is challenging to mix different types of material into a coherent search results display that will make sense to researchers. None of the research on discovery tools has evaluated the inclusion of OAI-PMH harvested material. The purpose of this study was to evaluate the usefulness of harvesting locally created collections of dissimilar content into Encore to enable their enhanced discovery. This study did not look at Digital Commons, which includes articles, or the Directory of Open Access Books (DOAB), which includes bibliographic materials, because those types of integrations have been studied elsewhere.

Harvesting for discovery case study

The interoperability of metadata is critical for successful harvesting. Standards for metadata consist of principles that address simplicity, modularity, reusability, extensibility, and interoperability. OAI is one tool that can be used to harvest metadata into a discovery tool. OAI-PMH has been in use for many years. It is one method of exposing metadata that can be collected or harvested to create a database for searching aggregated resources. It is a transport protocol that allows for the harvesting of metadata encoded in standard metadata schemes like Dublin Core (DC). The OAI standard is a simple protocol that uses the "get" and "post" HTTP requests with verbs that control the request actions. Timothy Cole and Muriel Foulonneau provided a comprehensive explanation of OAI in their 2007 book, *Using the Open Archives Initiative Protocol for Metadata Harvesting* (2007, 1–208).

OAI was developed to promote interoperability of content but leaves the work of metadata integration to the harvesting site. This task can be challenging for sources that do not have native support for OAI-PMH and require the library doing the harvest to develop a means for creating an output that can be used by an OAI aggregator. This process can include creating custom programs for transforming data that can be costly for a library.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

In the current age where "Google" has become a verb, the question arises, "Why bother"? After all, one of the biggest advantages of Google is the inclusive coverage of content from both popular and local sites as well as scholarly open source content. Peter Suber (2004, 1) made the case for OAI-compliant metadata. He argued that OAI makes research more visible than relying solely on Google for harvesting because librarians provide a quality control element that Google lacks. Suber asserted that scholars will turn to a trusted repository managed for research purposes before they will turn to Google. In addition, Suber noted that in comparison to Google, OAI provides standardized metadata schemas with the capability for specific field searching, more current index refreshing of content, and permanent URLs that are managed by repositories. Through selection of the "best" sources that include persistent URLs, standardized metadata fields, and quicker refreshment schedules, OAI-harvested searches can trump the short-term advantage of Google. Google is a search engine, not a repository, so it can only refresh the index to reflect what is currently available on the Web. Libraries that are creating repositories for long-term preservation should be concerned with the visibility of their data. One of the ways these data can be made more visible is through an OAI-compliant repository.

UNL implemented Encore in 2008 as a discovery tool. Encore is an Innovative Interfaces supported tool for integrating bibliographic records from the catalog with full-text resources (Allison 2010). In 2009, UNL began to integrate OAI-harvested metadata that uses several different metadata encoding schemes into Encore. The following OAI-harvested collections are currently available through Encore:

- Birds of Nebraska (http://birds-of-nebraska.unl.edu/), a Web site on birds in Nebraska gathered from newspaper articles and other sources from the years 1854–1923.
- Digital Commons (http://digitalcommons.unl.edu/), a Bepress repository of research articles contributed by UNL faculty and staff.
- Directory of Open Access Books (DOAB) (http://www.doabooks.org/), a full-text open access book repository.
- Elia Peattie (http://plainshumanities.unl.edu/peattie/), an archive based on the life and writings of Elia Peattie, an early Nebraska journalist, novelist, short story writer, poet, and playwright.
- Higginson (http://higginson.unl.edu/), the correspondence of Thomas Wentworth Higginson, an associate of Emily Dickinson.
- CONTENTdm (Image and Multimedia) (http://CONTENTdm.unl.edu/), a diverse collection of images provided by departments around the University that includes open and closed access.
- Lewis and Clark Journals (http://lewisandclarkjournals.unl.edu/), a Web site for the text of the Nebraska edition of the Lewis and Clark journals, edited by Gary E. Moulton.
- Subject and Course LibGuides (http://unl.libguides.com/), a collection of library guides to assist users in finding information.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

- University Archives Finding Aids (no Web site), a database containing descriptions of content in the University of Nebraska Archives.
- Willa Cather Archive (http://unl.libguides.com/), a Web site containing Willa Cather texts and scholarship open to the public with born-digital scholarly content.
- Performances from UNL's GKSOM (http://collections.unl.edu/GKSoMPerformances.html), a repository of streaming music and PDF programs from performances of UNL faculty and students, restricted to the UNL community.
- UNL Data Repository (https://dataregistry.unl.edu/), a database of research data contributed by UNL faculty and students.

The UNL Libraries uses OAI-PMH to harvest the metadata for locally created collections, which include image collections in CONTENTdm, articles, theses and dissertations in Digital Commons, and research data deposited in the UNL data repository. All of these data are incorporated into our discovery tool. In addition, other small specialty collections (both local and non-local that do not fall into one of the other three repositories) are also harvested.

Digital Commons, CONTENTdm, and DOAB all support OAI-PMH harvesting while the other small specialty collections did not have native support for OAI harvesting. Modified software was developed to expose locally created collections for harvesting. This modification consisted of extracting metadata from the source according to the OAI-PMH protocol that could then be harvested using UNL Libraries' designed XSL stylesheets and OAI.pl developed by Virginia Polytechnic Institute and State University. This extraction was done for Birds of Nebraska, Higginson, Lewis and Clark, Elia Peattie, the Willa Cather Archive, LibGuides, and the University Archives Finding Aids. These collections used metadata schemes including Text Encoding Initiative (TEI), Encoded Archival Description (EAD), and DC that were harvested for Encore searching.

The implementation of Encore at the UNL Libraries uses Synergy, which does not integrate articles into the results but presents a sample of articles in a preview box near the top of the results. This display choice has the advantage of not overwhelming patrons with articles but has the disadvantage of requiring an additional click to limit results to just articles. Because the box is not obvious, there is also the possibility that patrons will miss the articles entirely. In contrast, harvested items are integrated into the results (see Figure 1). The search displayed in Figure 1 demonstrates the integration of harvested materials and cataloged materials with a preview of article results for the search "welcome heroes." Images appear for some items because CONTENTdm consists principally of images, and when this collection is harvested an image appears with the metadata in the search results.

Encore supports an option to restrict searching to a particular collection by using the "Limit by Collection" facet in the navigation column. In addition, using the advanced search option, researchers can initially limit a search to a particular collection (see Figure 2).

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

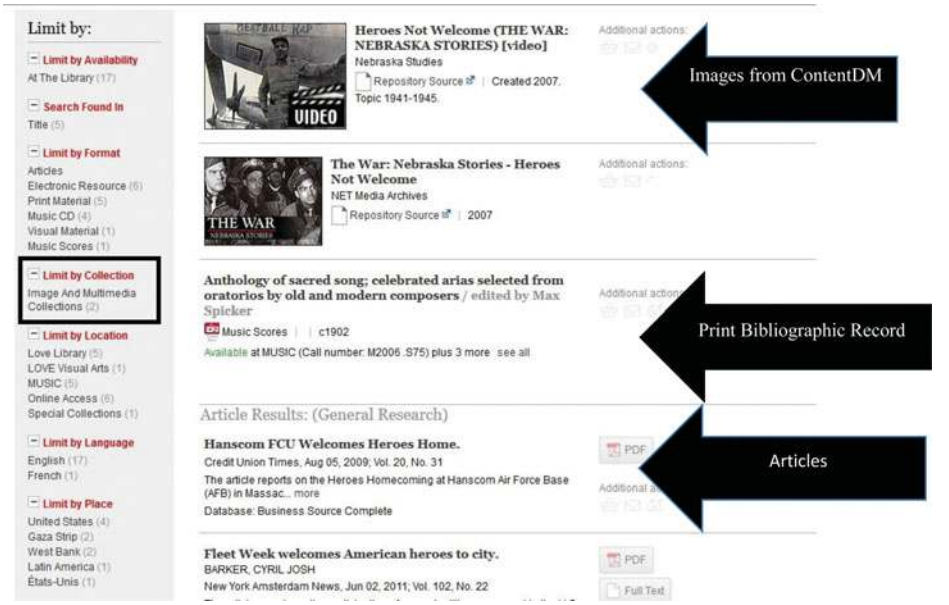


Figure 1. Integrated search results show harvested items grouped by collections.

Using these features, a researcher can search within a targeted, harvested collection. Each of these small collections contains appropriate metadata that describes the thematic content. For example, the UNL Libraries' instance of CONTENTdm consists of over 100 collections and sub-collections of images but is harvested together as a single collection. Each image within CONTENTdm is described with

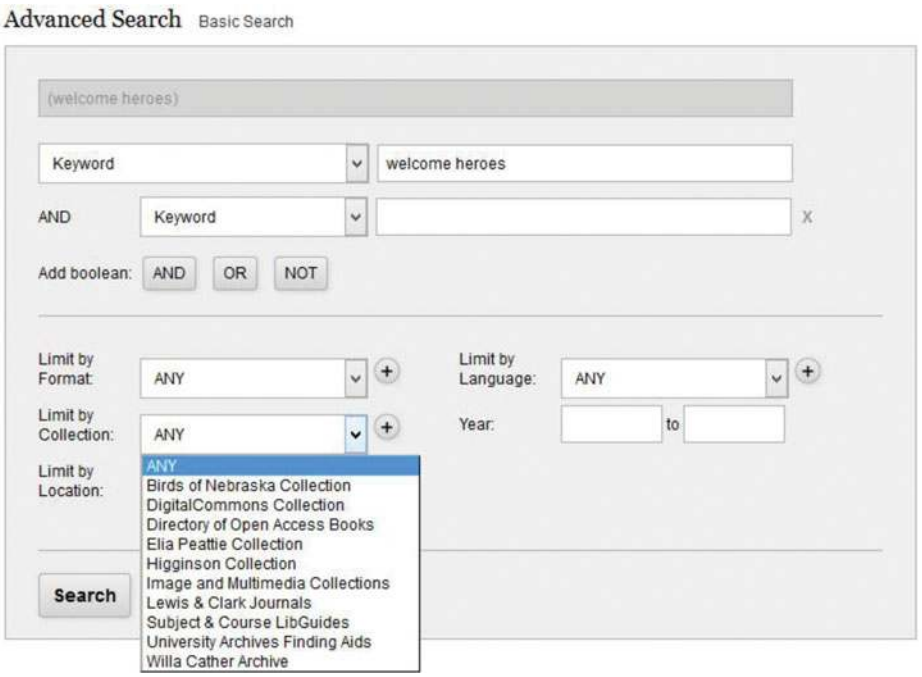


Figure 2. Advanced search option to limit search to a specific harvested collection.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

separate metadata that is searchable in Encore. The image is then displayed in the search results under the general collection heading of Multimedia and Image Collection.

Analysis of OAI-harvested content use

Google Analytics is a tool that can be used for collecting information on patron behavior. Huttenlock and Malone (2013, 366–85) discussed how Google Analytics can be used as part of an assessment strategy for gathering information that informs change in academic libraries. For example, at the 2015 Florida Library Association Conference, the Tampa-Hillsborough County Public Library reported on a project to redesign their Web site using information gathered using Google Analytics (Schane 2015, 211–13).

Google Analytics was selected for this study to identify differences in patron use of content discovered through OAI-harvested content and referrals from other origins. Although Google Analytics is an excellent source of information about patron behavior, it has the drawback of relying on users for "permission" to share information. Some people are adverse to this type of data collection and disable their sharing settings, so Google does not collect the information from those users. Nevertheless, the available data from patrons who allow their data to be collected can be analyzed as a reasonable sample. This study collected data from FY 2014/2015.

Google Analytics collects information based on the originating hostname or Internet Protocol (IP) number. The IP number consists of a series of digits divided by periods that trace a connection back to an individual computer. At UNL, IP numbers can either be static or dynamic. A static number never changes, whereas dynamic IP numbers are assigned on-the-fly when users connect to the network. In either case, IP numbers and hostnames are identified in Google Analytics by translating the numbers and hostnames into domains.

Google Analytics measures information by collecting data on the following:

- Sessions (a single connection that may include multiple events);
- Users (the number of new and returning users);
- Page views (the number of pages viewed, which includes counting repeated views of a page);
- Pages per session (the average number of pages in a session);
- Session duration (the average time for a session);
- Bounce rate (single page visits); and
- Percentage of new sessions (an estimate of the percentage of first-time visits).

Google Analytics also provides information on traffic sources. A referral occurs when someone clicks on a hyperlink on a Web page that connects to the site being tracked. Direct (sessions originating from within the site), organic (sessions originating from a search engine), and social sources that include social networks like Facebook, Twitter, and SlideShare, are also tracked. Google also offers a breakdown of session duration.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

In Google Analytics, there is a category for engagement that uses the length of time in a session as a measure of engagement, implying that longer viewing times are an indication that users are interested in the content. This assumes that session length is a qualitative indicator that people are spending time on the site for positive reasons. However, Google Analytics reports cannot tell us why people are looking at our pages and provide even less insight into how the information is being used. Nevertheless, the amount of time spent on a local collection is one indicator of usefulness even if it is not the whole story. For that reason, it was used in this study.

Another area of interest is the pattern of return visitors, which is considered a measure of loyalty, and for a non-profit, another factor of interest. Google Analytics creates a unique visitor ID that is stored in a "browser cookie" that tracks user activity (provided the user allows such tracking). When statistics are generated, data from any ID created before the reporting period, which reappears during the period under review, are counted as returning visitors.

To learn about user behavior for a range of the Libraries' OAI-harvested collections, this study was broken into two parts. The first investigation in this study uses Google Analytics data for the smaller harvested collections referred through Encore. The smaller harvested collections include metadata on locally created resources at the UNL Center for Digital Research in the Humanities (CDRH) that patrons may not associate with a library catalog. The CDRH is involved in creating databases and full-text resources on a variety of subjects and formats that include literary writers and historical figures and events (see the description of these collections in the case study section above). The objective for this first examination was to evaluate the usefulness of including dissimilar content as represented by the smaller collections as a subset of a much larger discovery tool.

The second part of the study focused on user behavior in CONTENTdm because it consists chiefly of images, which are very different from the articles in Synergy and monographs harvested from DOAB, and as such may result in different discovery trends. Using data from Google Analytics, the number of users, sessions, and duration were compared between Encore, CONTENTdm, and Encore referrals to CONTENTdm. This second analysis examined differences between user behaviors based on the referral source with the purpose of identifying possible trends in activity.

Analytics results

Statistics were gathered from Google Analytics based on Encore harvest content (see Table 1). Information for the data repository was not available at the time this article was written. In general, the larger the collection, the more likely it is to appear in any search results. This is most likely why there is a small number of users referred to these small collections. For example, the Higginson collection, which only has 38 records, did not appear in any search results. The Lewis and Clark Journals collection is so specialized that only a very targeted search would bring up results from this collection. The number of sessions (thirteen) compared with the return

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

Table 1. Google Analytics summary data for OAI-harvested collections.

Harvested collection	Number of records	Sessions	Users	Page views Page views	Avg. duration	Percentage of of new sessions	Percentage of return visitors
Lewis & Clark Journals	1,025	6	6	68	13:33	0	86
Birds of Nebraska: Newspaper Accounts, 1854–1923	1,023	13	12	231	16:50	31	56
Willa Cather Archive	467	8	7	113	7:00	50	67
University of Nebraska Archives Finding Aids	322	13	13	203	12:15	31	61
LibGuides	284	7	7	105	18:01	43	42
Elia Peattie: An Uncommon Writer, an Uncommon Woman	172	2	3	34	8:50	0	100
Writings of Thomas Wentworth Higginson	38	0	0	0	0:00	0	0

visitors percentage (56 percent) for the Birds of Nebraska is interesting and provides some evidence that people are finding and returning to the collection in spite of the speciality and small size of that collection. This pattern is repeated with the Willa Cather Archive (eight sessions and 67 percent return rate) and LibGuides (seven sessions and 42 percent return rate).

CONTENTdm collections are maintained principally by library staff, and many of the sub-collections were developed to support course content at UNL. To determine how users were getting to content in CONTENTdm, data were gathered from Google Analytics on the CONTENTdm site, which includes all traffic sources to CONTENTdm, as well as traffic referred from Encore. Figure 3 shows the majority of connections come from the .edu domain (51 percent). Most of the .edu searches are from within the university, which is not surprising since many of the collections were created for classroom support. Web/social sites (.com) that include a mixture of organic search engines, direct connections, and social sites (e.g., Facebook, Twitter, and Pinterest) are the second largest group (23 percent). The third category is

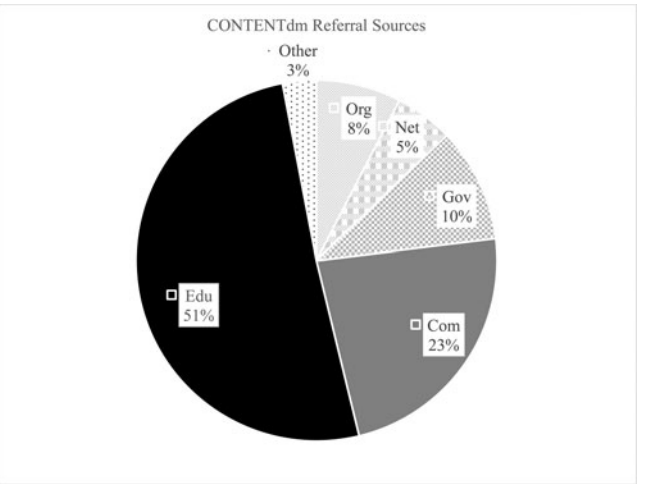


Figure 3. CONTENTdm referral sources by domain type.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

Table 2. Session comparisons between Encore, Encore referrals to CONTENTdm, and CONTENTdm.

Sept. 2014–Feb. 2015	Encore all activity totals	Encore referrals to CONTENTdm	CONTENTdm all source originations
Sessions	100,084	201	26,657
Users	30,333	142	21,539
Page views	472,922	3,313	94,201
Average pages per session	5	19	4
Average session duration	5:22	18:16	2:15
Bounce rate	37%	0%	64%
Percentage of new sessions	27%	40%	84%

the .gov domain (10 percent), which includes partners linking from their Web pages to shared digital collections.

Session statistics gathered from all sessions in Encore for all searches (including content from all harvested collections, the catalog, and articles), activity from Encore referrals to CONTENTdm, and all session activity in CONTENTdm, were compared (see Table 2). The average pages per session were considerably higher for connections referred through Encore (nineteen) than for the other sources. Likewise, the session duration for those sources referred through Encore averaged 18 minutes compared to 2 and 5 minutes that are typical for other sources. The percentage (40 percent) of new sessions referred to CONTENTdm through Encore is higher than sessions initiated in Encore (27 percent) but lower than sessions initiated via all other sources to CONTENTdm (84 percent). The majority of new sessions referred to CONTENTdm come through a search engine or click-throughs from links in a Web site. The percentage of new sessions referred through Encore is lower than referrals to CONTENTdm through other means, which is logical given the small total numbers referred through Encore.

When using Google Analytics to examine the amount of time users spend on a page, the statistics show that when sessions are referred from Encore, users spend more time looking at content than when they are referred from other connections (see Table 3). While 23.7 percent of those coming through Encore spent 30 minutes or more on images, only 1.4 percent of those coming from other sources spent that much time.

Repeat visitor statistics revealed that Encore referrals to CONTENTdm have an overall visitor return rate of 55.6 percent, with 18.7 percent returning nine times or more (see Table 4). Although the total number of returns to CONTENTdm from all

Table 3. Engagement with CONTENTdm measured by time spent per user session.

Engagement in seconds	Encore referral	Percent of users referred through Encore	All connection sources in CONTENTdm	CONTENTdm time %
0–10	3	1.7%	16,838	66.2%
11–30	13	7.9%	1,582	6.2%
31–60	12	6.8%	1,415	5.6%
61–180	24	13.6%	2,116	8.3%
181–600	38	21.5%	1,856	7.3%
601–1,800	45	25.4%	1,256	5%
1,800+	42	23.7%	368	1.4%

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

Table 4. Repeat visitors by referral from Encore and other sources.

Number of sessions	Encore referral: Number of users	Encore referral percentage	CONTENTdm: Number of users	CONTENTdm percentage
1	71	44.4%	21,368	84.0%
2	17	10.6%	2,074	8.1%
3	15	9.4%	617	2.4%
4	5	3.1%	294	1.1%
5	8	5.0%	188	0.7%
6	8	5.0%	126	0.5%
7	2	1.3%	92	0.3%
8	4	2.5%	72	0.2%
9–14	13	8.1%	252	0.9%
15–25	12	7.5%	185	0.7%
26+	5	3.1%	160	0.6%

other sources is higher, the overall percentage of returns for nine visits or more is only 2.2 percent.

Discussion

There were several findings from this study, including the following: A small number of referrals were directed to harvested sites in comparison with total site activity; the amount of time spent on the harvested sites was higher for referrals through Encore than when referrals came from other sources; and Encore referrals had a higher percentage of visitor return than other types of referrals. The overall visits to harvested material from the CONTENTdm collection were small when compared with the numbers coming through other discovery means (see Table 2). This is in contrast with other research that demonstrates increased use of journals when full-text articles were added to a discovery tool (Levine-Clark et al. 2014). This finding is not surprising considering the specialized content of the harvested collections. For example, the CONTENTdm collection is diverse and covers content from art, architecture, and the sciences, but it is highly specialized in that it includes images of individual items, which would require precise search terms to match.

The average session duration for all OAI-harvested collections is impressive since records for collections referred through Encore exceeded the overall average session time of five minutes that is generally spent searching Encore (see Table 1). In the case of collections in CONTENTdm, differences in the average session time could be influenced by the types of content in the collection (see Table 2). Since CONTENTdm is used by classes, it may be that the majority of direct connections are generated by instructors directing students to a single image, whereas the Encore referrals are more exploratory.

Recommendations differ about how to evaluate statistics for visitor returns. The numbers for visitor returns will vary based on the type of Web site, so recommendations for "good numbers" vary from 4 percent to 20 percent. Seb Chan (2015, 1) reported that social sites have a return rate from a low of 12 percent for Wikipedia to a high of 42 percent for Flickr, with Facebook at 33 percent, Twitter at 34 percent, and the average for all of these social sites at 21 percent. Given the small size of the

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

UNL collections, a visitor return rate of 25 percent is evidence for a successful rate of return. All the smaller collections (see Table 1) except Higginson (0 percent) and LibGuides (42 percent) have return rates higher than 50 percent. Higginson had no activity during the study period, and the LibGuides collection was a new collection with fewer records at the time of this study. With the beginning of the fall term, many new LibGuides were added to the collection that are used by librarians in instruction and reference sessions. For the month of October 2015, the return visitor rate for LibGuides was 73.5 percent, so it appears that the return visitor rate may be influenced by the growth rate of a dynamic collection. Further research may be able to determine what influence the size and vigor of a collection has on user behavior.

The longer amount of time spent exploring images, the low bounce rate, and higher percentages of referred users coupled with a high return rate for CONTENTdm searches poses an interesting but cautionary theory about library discovery tools. The quality of library content may be attracting users to Encore who then spend more time on content than they would if they accessed the content through a Web search. The reason for higher engagement numbers for the OAI-harvested content is an interesting question to explore in further research.

Conclusion

What is a reasonable return on investment for harvesting? There were over 169,000 sessions in Encore during FY 2014/15 with only 834 sessions referred to any of the harvested collections. Users spent an average of 5 minutes looking at items that were in Encore while the overall average for referrals to harvested collections was 15 minutes. According to an article in *Time* by Tony Haile (2014, 1), the length of time a reader spends on a Web page is one of the best indicators for judging a successful page. He indicated that 55 percent of readers spend 15 seconds on a Web page, so a visit of 15 minutes can be considered one measure of success.

This study did not address the question of how the content discovered through Encore is being used or the significance of the content to patron research. The prominent placement of the Encore search box on the Libraries' Web site will draw users to Encore. Some instructors are referring their students to go directly to one of the harvested sites instead of using Encore for discovery. Other patrons use a search engine to get to content. This study only evaluated the indicators of behavior between OAI content discovered in Encore versus direct connections and referrals. Additional research is needed to replicate these findings and to provide more light on why there are differences.

Based on this research, libraries considering adding OAI-harvested items into their discovery tool should not expect their harvest sites numbers to increase dramatically. Individual results will vary based on the collections harvested and marketing to target audiences. Once harvests are discovered, libraries can expect the use of these collections to increase as measured by the number of pages viewed, return

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

visits, and time spent exploring content. The high return visitor rates found in this study provide some evidence for Suber's (2004) argument that scholars are looking for the quality that libraries using OAI harvesting provide.

Libraries are in the business of assisting patrons one at a time. Given the amount of resources devoted to collecting and assembling local collections, a discovery tool that includes these resources appears to be a reasonable investment for our users. Harvesting OAI sources into discovery tools seems to increase the value of a discovery tool for scholars who rely on libraries to provide the highest quality research material.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016

About the author

DeeAnn Allison is the Director for Computing Operations & Research Services and a professor at University of Nebraska-Lincoln. She oversees the Libraries ILS, Encore discovery tool, and she coordinates technology with other library faculty and departments at the University. Her research interests focus on finding methods for evaluating the use of library resources and ways libraries are influencing research.

References

Allison, D. 2010. Information portals: The next generation catalog. *Journal of Web Librarianship* 4 (4):375–89.

Asher, A. D., L. M. Duke, and S. Wilson. 2013. Paths of discovery: Comparing the search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and conventional library resources. *College & Research Libraries* 74 (5):464–88.

Breeding, M. 2015. The future of library resource discovery: A white paper missioned by the NISO discovery (D2D) topic committee. http://www.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf (accessed June, 2015).

Bull, S., E. Craft, and A. Dodds. 2014. Evaluation of a resource discovery service: FindIt@Bham. *New Review of Academic Librarianship* 20 (2):137–66. doi:10.1080/13614533.2014.897238.

Calvert, K. 2015. Maximizing academic library collections: Measuring changes in use patterns owing to EBSCO Discovery Service. *College & Research Libraries* 76 (1):81–99. doi:10.5860/crl.76.1.81.

Chan, S. 2015. Which social web platforms create the most return visitors to our website? Fresh & new(er) discussion of issues around digital media and museums. http://www.freshandnew.org/2010/09/which-social-web-platforms-create-the-most-return-visitors-to-a-website/(accessed June , 2015).

Cole, T. W., and M. Foulonneau. 2007. *Using the open archives initiative protocol for metadata harvesting*. Westport, CT: Libraries Unlimited.

Condit Fagan, J., M. Mandernach, C. S. Nelson, J. R. Paulo, and G. Saunders. 2012. Usability test results for a discovery tool in an academic library. *Information Technology & Libraries* 31 (1):83–112.

Djenno, M., G. Insua, G. M. Gregory, and J. S. Brantley. 2014. Discovering usability: Comparing two discovery systems at one academic library. *Journal of Web Librarianship* 8 (3):263–85. doi:10.1080/19322909.2014.933690.

Haile, T. 2014. What you think you know about the web is wrong. *Time Com*: 1–1. http://time.com/12933/what-you-think-you-know-about-the-web-is-wrong/ (accessed Jan. 19, 2016).

Huttenlock, T., and D. B. Malone. Proving value and improving practice: Using system data to analyze user behaviors. *College & Undergraduate Libraries* 20 (3/4):366–85.

Johns-Smith, S. 2012. Evaluation and implementation of a discovery tool. *Kansas Library Association College and University Libraries Section Proceedings* 2 (1):17–23.

Levine-Clark, M., J. McDonald, and J. S. Price. 2014. The effect of discovery systems on online journal usage: A longitudinal study. *Insights: The UKSG Journal* 27 (3):249–56. doi:10.1629/2048-7754.153.

Schane, S. 2015. Florida Library Association Annual Conference Report: The user focused online library. *Journal of Electronic Resources Librarianship* 27 (3):211–13.

Suber, P. 2004. The case for OAI in the age of Google. *SPARC Open Access Newsletter* 73 (May 3):1. http://dash.harvard.edu/bitstream/handle/1/4455494/suber_googleoai.htm?sequence=1.

Thomsett-Scott, B., and P. E. Reese 2012. Academic libraries and discovery tools: A survey of the literature. *College & Undergraduate Libraries* 19 (2–4):123–43. doi:10.1080/10691316.2012.697009.

Downloaded by [University of Nebraska, Lincoln] at 04:44 17 February 2016