Revisiting the cataloging of free Internet resources at the University of Denver

Christopher C. Brown and Erin Elzi
University Libraries, University of Denver, Denver, Colorado, USA

Abstract

Purpose – This paper aims to present updated statistics demonstrating the value of cataloging free Internet resources and the challenges of batch loading, vendor records, electronic resource modules and discovery tools, as an update to the 2008 paper in this journal

Design/methodology/approach – Updates the statistics from the URL redirection system for tracking user access to freely available Web publications.

Findings – With more projects and bibliographic records included within the scope of the project, users still find and use the links to outbound content. New technologies and management methods support the cataloging of free Web content, even if, at times, cataloging standards are compromised.

Originality/value – Several studies have focused on US federal document clickthroughs from the library catalog, but this is the only study to exclusively track clickthroughs to freely available Web content.

Keywords User studies, Library users, Electronic document delivery, Academic libraries, Cataloguing, Collections management

Paper type Case study

Preamble

In 2008, Brown and Meagher published "Cataloging free e-resources: is it worth the investment?" in this journal (Brown and Meagher, 2008). The authors sought to determine if was worth the time and effort to catalog freely available Internet resources. This project was an added dimension to a project already begun, in which they tracked user clickthroughs to US federal publications, a project documented in a separate paper (Brown, 2011). They tracked the usage of these non-federal free resources by adding clickthrough links to the 856 fields in catalog records.

The University of Denver (DU) library has continued to maintain this procedure for freely available Web content. This follow-up article seeks to answer two questions:

Q1. Is the original premise true that cataloging free Internet resources is worth the investment?

Q2. What has changed in the intervening years that affects the ways we provide access to free Internet resources?

Background to the study

Tracking of outbound links from the online public access catalog was begun by DU in 2003 because an increasing number of URLs were appearing in catalog records for government publications, and nothing analogous to vendor-supplied clickthrough data was available for these resources. When the project was extended to other free Internet resources in 2004, the same lack of statistics motivated these data gathering efforts. The arrival of Google Analytics in 2005 could not lend assistance to this project, as the links were to content outside the library's purview, and thus not able to be tracked.

The clickthrough mechanism used at that time was a URL prefix, appended before the free resources' URL, as seen in Figure 1. When a user clicked the URL in a catalog record, the URL first went to the project site, where the URL, date/time, project abbreviation code and IP address of the user's computer were logged (as seen in Table I), following which the user was redirected to the desired Web URL.

The DU Library was one of the first libraries to track URL clickthroughs by placing redirects in the library catalog. From 2003, the Library focused exclusively on US federal government publications. About one year later, the Library folded into the project other freely available Internet content, going beyond US Government publications to the State of Colorado publications, intergovernmental organizations, non-governmental organizations and selected publishers with free content. In 2009, representatives from six libraries across the USA gave a presentation at the annual Federal Depository Library Conference in which they shared their click-tracking technologies and the results of such tracking (Brown, 2009). It became clear that no two libraries implemented clickthrough tracking in exactly the same manner, and that this tracking provided significant collection development guidance to libraries seeking to provide relevant resources to their populations.

Although at least five libraries are involved in locally tracking federal government information through clickthrough

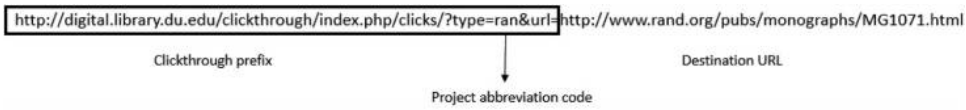
The current issue and full text archive of this journal is available on Emerald Insight at: **www.emeraldinsight.com/0264-1615.htm**

Interlending & Document Supply
44/1 (2016) 31–36

[DOI 10.1108/ILDS-11-2015-0034]

Received 9 November 2015
Accepted 10 December 2015

Figure 1 Clickthrough prefix appended before destination URL



Note: The project abbreviation code "ran" – for RAND – in the example

Table I Clickthrough transaction log

Date	Time	Type	IP	Url
6/23/2015	15:18:02	FRO	162.244.12.133	www.pbs.org/wgbh/pages/frontline/topsecretamerica/
6/23/2015	15:32:22	ERI	50.168.200.248	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED546463
6/23/2015	15:49:17	RAN	10.8.32.26	www.rand.org/publications/MR/MR1198/
6/23/2015	16:47:00	ERI	65.114.255.190	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED415385
6/23/2015	17:59:05	ERI	10.8.36.41	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED492959
6/23/2015	19:07:26	ERI	209.181.64.229	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED525059
6/23/2015	19:11:03	GOV	130.253.30.18	www.osti.gov/servlets/purl/894551-EXEiEl/
6/23/2015	19:11:56	ERI	209.181.64.229	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED525056
6/23/2015	20:06:18	ERI	209.181.64.229	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED538604
6/23/2015	20:26:55	GOV	50.183.230.11	http://purl.access.gpo.gov/GPO/LPS93604
6/23/2015	21:24:53	COU	73.14.89.124	http://hdl.handle.net/10176/co:5381_ucb610916bpm112007internet.pdf
6/23/2015	21:25:43	COU	73.14.89.124	http://hdl.handle.net/10176/co:5381_ucb610916bpm112007internet.pdf
6/23/2015	21:26:32	GOV	73.14.89.124	http://purl.fdlp.gov/GPO/gpo52147
6/23/2015	21:26:48	ERI	71.218.130.190	www.eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED537714

projects, very few projects for other freely available content are known to these authors.

Since the 2008 article, ERIC (Education Resources Information Center) MARC records became available from the vendor Marcive. Originally, the entire set of over 300,000 ERIC records were loaded into the DU catalog. Liza Weisbrod, librarian at Auburn University, also oversaw a clickthrough project for freely available Web content via the Auburn library catalog. After loading ERIC MARC records, she noted that 27 per cent of all links in the library catalog were links to ERIC content, and that 5 per cent of the clickthroughs were to ERIC content (Weisbrod, 2011).

But on August 3, 2012 the entire set of ERIC documents PDFs were taken offline, due to the discovery of personally identifiable information in some of the documents (Pollard, 2014). When users became frustrated with the link failure in these documents, the MARC records were all suppressed from the DU library catalog. As ERIC restored selected PDFs, the records were gradually made available again to users.

Update to the original study

Since the 2008 article, the DU library catalog has changed from a library-hosted system to one hosted by the DU's University Technology Services. In 2010, the original clickthrough tracking mechanism had to be migrated from a ColdFusion platform and the code completely rewritten in PHP language. In 2011, the Library upgraded from Innovative Interface's Millennium platform to the Sierra platform. Yet throughout this entire period we have continued to track user clicks of outbound links in catalog records to freely available Internet content. The technology on the PHP platform remains the same conceptually as it was when originally designed.

Overview of projects

Many of the projects (listed in Table II) were in place at the beginning of the clickthrough project and still are in place to this day. These include American Museum of Natural History, American Mathematical Society, Brookings Institution, Council on Library and Information Resources, Colorado State Publications, Documenting the American South, Human Rights Watch, Institute of Development Studies (UK), Making of America, National Academies Press, University of California Press, United Nations, Wright American Fiction and the World Bank. As time went by, several other sets of freely available Internet sites were added to the project, the most significant of these being ERIC.

There are many variables to consider, some difficult to test, when it comes to summarizing our statistics. By 2014, we had changed our library discovery experience. No longer was the library catalog the only prominent search box; our Summon (ProQuest) discovery tool was the most prominent box, with our catalog being secondary. Even though catalog records are loaded into Summon, records for free information and clickthroughs are in a more diffuse environment than the catalog, having a backgrounding effect. These changes were documented in the study conducted by Brown (2013). Even though the library Web site underwent several design changes, users were generally still able to benefit from the time and expense of adding the records for freely available resources.

The 2008 article documented the number of bibliographic records in the catalog for each of the tracked projects. Also shown was the per cent of these records that had been clicked at least one time. Table III restates the 2007 statistics along with the records for selected projects from 2014. This table shows the number of catalog records for selected projects at two snapshots in time: 2007 and 2014. This differs from the

Table II Clickthrough totals of selected projects, 2005-2007 and 2012-2014

Project name	Abbreviation	Clicks 2005	Clicks 2006	Clicks 2007	Unique clicks, 2005-2007	Clicks 2012	Clicks 2013	Clicks 2014	Unique clicks, 2012-2014
American Museum of Natural History	AMN	0	1	3	4	2	0	8	6
American Mathematical Society	AMS	1	5	1	6	3	0	0	3
Brookings Institution	BRO	14	5	7	16	1	3	16	5
Council on Library and Information Resources	CLR	42	41	12	43	16	2	61	16
Colorado State Publications	COU	12	189	561	388	456	424	591	883
Documenting the American South	DAS	14	18	13	28	3	5	8	12
ERIC Documents	ERI	0	0	0	0	2,699	2,341	4,571	5,991
Frontline	FRO	0	0	0	0	27	38	724	82
Google Books	GOO	0	0	0	0	4	3	6	11
Project Guttenberg	GUT	0	0	0	0	111	85	140	158
HathiTrust (non-docs)	HAT	0	0	0	0	9	9	6	12
Human Rights Watch	HRW	32	36	18	47	13	18	34	38
Institute of Development Studies (UK)	IDS	0	8	27	21	19	5	16	21
International Labor Organization	ILO	0	0	0	0	3	1	6	6
International Organization for Migration	IOM	0	0	0	0	8	6	3	9
Lonely Planet	LPT	0	0	0	0	157	256	480	92
DU Law CO Legislative Council	LRC	0	0	0	0	13	6	8	20
Macbeth Gallery	MAC	0	0	0	0	4	21	10	25
Making of America	MOA	289	255	217	516	84	50	112	185
Miscellaneous	MSC	331	636	538	492	2,706	725	521	454
National Academies Press	NAP	246	268	322	410	910	287	191	455
Public Broadcasting System	PBS	0	0	0	0	0	0	3	2
Posner Library (Carnegie Mellon)	POS	0	0	0	0	7	7	17	19
Rand	RAN	143	169	171	324	294	148	127	297
University of California Press	UCP	14	95	53	109	21	22	35	46
United Nations	UNN	79	119	149	163	391	273	145	136
Wright American Fiction	WAF	44	40	46	102	14	12	24	35
World Bank	WBK	20	38	14	7	12	6	5	6
Totals		1,281	1,923	2,152	2,676	7,987	4,753	7,868	9,025

number of clickthroughs registered. Some projects may have very few catalog records (Table III), but many clicks (Table II). Other projects may have a great many catalog records, but fewer clicks per record. The wide swings in per cent accessed between 2007 and 2014 are difficult to explain, as there are multiple variables such as the addition of a discovery layer (Summon) with catalog records included and changes in user interface (prominence of catalog search box in 2007 and prominence of Summon in 2014). Further change is planned in 2016, as the Library transitions to the ExLibris Alma integrated library system with a Primo discovery layer.

Changes to record management and effects on record tracking

Since 2008, the availability of vendor-provided record sets for freely available resources has proliferated. Some of these MARC record sets are themselves freely available on the Web (Wright American Fiction, 2007), some require a login or have to be requested from the database host [e.g. Making of America (MOA MARC Records, 2006)], and others still are sets that can be purchased from the database host of an outside vendor [e.g. ERIC (Marcive Inc., 2015)].

As the workflows of many cataloging departments have seen an increase in the number of batch-loaded record sets and a decrease in the creation/enhancement of individual MARC records, incorporating the record sets for freely available resources requires a minimal addition to workloads. However, the continued management of these records poses problems. In the case of the ERIC records in 2012, fortunately it was a group of records that were easy to suppress and monitor. However, there is always the chance that the free resources records are not so easy to track. Nor is it feasible to regularly check all links to freely available content for stability.

As libraries have not purchased access to the content, the vendor or host may or not feel obligated to provide updates in a timely fashion. In some cases, if a database is no longer being updated or actively maintained, the host themselves might not even be aware of the lapse in access to individual records. It

Interlending & Document Supply
Volume 44 · Number 1 · 2016 · 31–36

Table III Overall comparison of selected clickthrough projects, 2007 and 2014

Project name	Abbreviation	Cat records 2007	Cat records 2014	% accessed, 2005-2007	% accessed, 2012-2014
American Museum of Natural History	AMN	69	70	5.8	11.7
American Mathematical Society	AMS	31	31	19.3	10.3
Brookings Institution	BRO	28	29	57.1	5.8
Council on Library and Information Resources	CLR	101	102	42.5	6.4
Colorado State Publications	COU	4,522	12,038	8.5	13.6
Documenting the American South	DAS	145	147	19.3	12.3
ERIC Documents	ERI	0	41,909	n/a	7.0
Frontline	FRO	0	102	n/a	1.2
Google Books	GOO	0	342	n/a	31.1
Project Guttenberg	GUT	0	488	n/a	3.1
HathiTrust (non-docs)	HAT	0	32	n/a	2.7
Human Rights Watch	HRW	188	192	25.0	5.1
Institute of Development Studies (UK)	IDS	104	124	20.2	5.9
International Labor Organization	ILO	0	25	n/a	4.2
International Organization for Migration	IOM	0	37	n/a	4.1
Lonely Planet	LPT	0	132	n/a	1.4
DU Law CO Legislative Council	LRC	0	301	n/a	15.1
Macbeth Gallery	MAC	0	451	n/a	18.0
Making of America	MOA	5,996	5,999	8.6	32.4
Miscellaneous	MSC	633	779	77.7	1.7
National Academies Press	NAP	1,989	2,772	20.6	6.1
Public Broadcasting System	PBS	0	6	n/a	3.0
Posner Library (Carnegie Mellon)	POS	0	112	n/a	5.9
Rand	RAN	1,225	2,096	26.5	7.1
University of California Press	UCP	387	389	28.2	8.5
United Nations	UNN	447	905	36.5	6.7
Wright American Fiction	WAF	2,839	2,841	3.6	81.2
World Bank	WBK	8	16	87.5	2.7
Totals/average		18,712	72,467	30.4	11.2

also becomes an issue as to who should be responsible for knowing if a freely available resource goes offline – should it be the cataloging department, acquisitions department, or the librarian, faculty member or collection development group that requested the records be added? One of the strengths of this clickthrough project is that we can monitor clickthroughs and periodically check selected links to ensure that the outbound links still work.

Electronic resource management and access subscription systems, such as ProQuest's Serials Solutions and EBSCONET Subscription Management systems, have made the management and loading of some record sets easier. The records are included with a regularly scheduled load that is already occurring for paid resource MARC records, and the management of stable links is done by the provider. Should the records need to be deleted from the system, this is also done with the regularly scheduled batch process for all of the electronic resource access system. Of course, no system is perfect. There can be delays between when a link changes or a resource is taken off-line and when the regularly scheduled load that contains these updates occurs. As the electronic resource access system acts as a middle man between the database or vendor and the library, changes made by the vendor may not be reflected in the catalog in a timely fashion, and sometimes even requires that a library approach the access system provider to request an update. Yet despite these drawbacks, the overall management of freely available resources has been made easier by these electronic resource access systems.

Management of the records aside, the quality of MARC records is an ongoing issue. Batch-loaded MARC records vastly vary in quality based on the provider, for both free-resources and purchased content. In some cases, there are various places where records are made available, and the choice to balance monetary cost, the number of records and their quality must be considered in a cost-benefit analysis. For example, the DU Libraries pay per record for ERIC records from Marcive and have received over 300,000 document-level records. The ERIC records from Marcive do not fully conform to Anglo-American Cataloguing Rules (AACR2) or Resource Description and Access (RDA) standards and ERIC-specific subject headings are used, which many integrated library systems (ILS) may not be configured to index. Even if an ILS is configured to index these ERIC-specific terms as subjects, they would not have the same functionality as subjects that have an authority record in the system. Despite these drawbacks, the descriptive cataloging is correct and the number of subject terms applied to each ERIC document is impressive.

Unfortunately, the same cannot be said of a significant number of the available record sets. The mediocre state of batch-received MARC records is an issue that occurs from all

Interlending & Document Supply
Volume 44 · Number 1 · 2016 · 31–36

types of sources – vendors, publishers, OCLC Worldshare, etc., forcing libraries to make the choice between quality (records) and quantity (access to resources). Previous studies have shown that in batch-loading MARC records for certain collections, the free records made available for e-books required a considerable number of corrections after record loading (Mugridge and Edmunds, 2012). In the case of free-content, it raises the question of whether we should be adding low-quality records for sources that can be found on the Web.

Many ILS systems now include an Electronic Resource Management system (ERM) that may require ERM records, bibliographic records and/or item-level records to be created for all database-level resources. Depending on the library's system and where the MARC records come from, these can be managed individually or as a batch process. Either way, this is another aspect of managing the free internet resources that needs to be considered and incorporated into workflows. In addition to the purchased and subscribed content managed through the ERM and Serials Solutions, DU uses Serials Solutions to bring in the MARC records for 41 free resources, plus seven packages of free journals that are compiled by Serials Solutions (ProQuest), for a total of 46,178 journals and e-books (as of October, 2015). Among these are the Directory of Open Access Journals (10,518), Galica Periodicals (11,050) and Making of America (8,503). The advantage to such a system, particularly when dealing with large packages of thousands of records, is the regularly scheduled receipt of updates from the record provider, something that does not happen when the publisher simply posts the free MARC records on their Web site. The drawback, again, is quality control. At the DU Libraries, the current ERM system has been set-up so that all records are overlaid each month. This makes the quality management of the records nearly impossible, as any improvements to records will be eliminated with the next load. In these cases, one must rely on the publisher, vendor or other record creator to make record improvements.

It is fair to say that the current state of e-resources is dominated by batch loads. However, freely available resources that are not part of a package are still plentiful. Among these are resources that are born digital, as well as those that are an alternative format to print material. It is far easier to exercise control over these records, as, like traditional cataloging for print material, the cataloger brings each record into the system one by one. The trade-off is the amount of time that goes into the process. Thousands of records can be batch loaded in the time it takes to bring a single record into the ILS, and that is assuming there is an existing record available, and that it meets the library's quality standards for copy-cataloging. All these for a resource that presumably could be found through a Google search.

There are further aspects to take into consideration when choosing to bring free e-resources into the catalog on a singular basis. One is whether or not the library has the print version. The issue here is twofold. First, should the online version be bothered with if the library holds the print copy? Then, if the answer to the first question is yes, how should the e-resource be treated in respect to the print copy? There are many ways to do this and consistency is the key. CONSER, the Cooperative Serials branch of the Library of Congress' Program for Cooperative Cataloging, allows for simply adding the link to the bibliographic record or attaching an additional item record to the print-version bibliographic record (for serials only), but AACR2 requires bringing in a separate bibliographic record for the e-resource (CONSER, 2013). Whereas AACR2 allowed for the use of the e-resource record to be derived from the print record (with some added elements), RDA goes a step further and requires that the e-resource be cataloged as its own entity, not as a reproduction or additional format of the print (PCC Task Group on RDA and Provider-Neutral Model and Reproductions, 2013). However, the needs of the catalog users and the configuration of the ILS are also a factor. DU decided in 2001 that when applicable, it would add an Internet item record to its print bibliographic record. Despite breaking the rules of national cataloging standards, DU users found that catalog search results were easier to navigate when the print and electronic version were on a single record.

Despite the additional workload to add resources that are already freely available on the Internet and the difficulties that their management causes, the ultimate factor in determining if cataloging these resources is worth it is the pay-off to the user. According to Beall's (2009) study of loading Google Scholar records, it is worth the extra effort to add these resources to the catalog, as the titles fill in gaps in the collection. In their 2012 study, Hill and Bossaller (2013) determined that the main factor prohibiting the successful use of free resources in the library catalog is not the cataloging effort required, but the lack of institutional policies regarding such resources – that it is done in a manner too haphazard for the researcher to feel confident in knowing they can find a wide and deep variety of free resources in the catalog. Cataloging free resources is worth it, but only if done consistently, according to a well-developed institutional policy and the effort is made to find the best records possible.

Conclusion

For more than 10 years, DU has been adding URLs to its local catalog for freely available Internet content. The tracking protocols used for these 10 years have shown that users do use this online content. Based on these statistics, the DU Library has concluded that the time it takes to catalog these freely available resources or to add links to existing records for tangible materials is indeed worth the effort.

Automation processes have changed the way that records can be acquired and loaded, and the workflows in the DU cataloging department have adjusted accordingly. Cataloging standards have been stretched to accommodate these efforts and vendor record loads have added to the project for freely available resources. These records are easier to deal with, but present challenges in terms of record quality. Creating workflows for these records and adapting to the new processes is what consumes the greatest amount of time in the department, but once they are established, the control of the records grows easier each year as more and more records are batch managed. As a result, the cost-benefit consideration leans in favor of adding the freely available resources to the catalog: a vast number of resources are added with little extra effort, therefore only a minimal amount of usage by researchers is required to make the tasks taken by the cataloging department worthwhile. The records added to the DU catalog are carefully chosen sets that,

despite the difficulties in record management, do fill in gaps and therefore provide added value to the researchers.

References

Beall, J. (2009), "Free books: loading brief MARC records for open-access books in an academic library catalog", *Cataloging & Classification Quarterly*, Vol. 47 No. 5, pp. 452-463.

Brown, C.C. (2009), "Tracking online document usage from the catalog: experiences from the field", presentation given with Stephanie Braunstein, Susan Kendall, Liza Weisbrod, Jennifer Gerke, and Shane Cole at the Fall 2009 Depository Library Conference, 19 October, Arlington, VA.

Brown, C.C. (2011), "Knowing where they went: six years of online access statistics via the online catalog for federal government information", *College & Research Libraries*, Vol. 72 No. 1, pp. 43-61.

Brown, C.C. (2013), "Web analytics applied to online catalog usage at the university of Denver", in McHale, N.L. and Farney, T.A. (Eds), *Web Analytics Strategies for Information Professionals: A LITA Guide*, ALA Editions, Chicago, IL, pp. 167-181.

Brown, C.C. and Meagher, E.S. (2008), "Cataloging free e-resources: is it worth the investment?", *Interlending & Document Supply* , Vol. 36 No. 3, pp. 135-141.

CONSER (2013), *CCM Module 31, Remote Access Electronic Serials*, available at: www.loc.gov/aba/pcc/conser/word/Module31.doc (accessed 30 October 2015).

Hill, H. and Bossaller, J. (2013), "Public library use of free e-resources", *Journal of Librarianship and Information Science*, Vol. 45 No. 2, pp. 103-112.

Marcive, Inc (2015), "ERIC MARC Records", available at: http://home.marcive.com/ericbackfile (accessed 24 October 2015).

MOA MARC Records (2006), available at: http://quod.lib.umich.edu/m/moagrp/moa-marc.html (accessed 24 October 2015).

Mugridge, R.L. and Edmunds, J. (2012), "Batch loading MARC bibliographic records: current practices and future challenges in large research libraries", *Library Resources & Technical Services*, Vol. 56 No. 3, pp. 155-170.

PCC Task Group on RDA and Provider-Neutral Model and Reproductions. (2013), *Provider- Neutral E-Resource MARC Record Guide: P-N/RDA Version*, 1 January, available at: www.loc.gov/aba/pcc/scs/documents/PN-RDA-Combined.docx (accessed 30 October 2015).

Pollard, E. (2014), "Restoring access to ERIC's PDFs", available at: http://eric.ed.gov/pdf/ERIC_PDF_Restoration_Webinar.pdf (accessed 12 October 2015).

Weisbrod, L. (2011), "Bringing ERIC documents to life in Voyager", Auburn University, available at: http://home.marcive.com/wp-content/uploads/2013/09/ERICcaseAuburn2.pdf (accessed 11 October 2015).

Wright American Fiction (2007), "Wright American Fiction 1851-1875", available at: www.tei-c.org/Activities/Projects/wr01.xml (accessed 24 October 2015).

Further reading

Guay, B., Rachel, B.S. and King, D. (2014), "Management of e-resources cataloging workflows at the University of Maryland, College Park", *Journal of Library Innovation*, Vol. 5 No. 2, pp. 29-51.

Corresponding author

Christopher C. Brown can be contacted at: cbrown@du.edu