

Digital Collections Usage at the University of Illinois at Urbana- Champaign Library

2015 REPORT

KYLE R. RIMKUS AND KELLY J. APPLGATE

Digital Collections Usage at the University of Illinois at Urbana-Champaign Library: 2015 Report

Kyle R. Rimkus

Preservation Librarian and Assistant Professor
University of Illinois at Urbana-Champaign

Kelly J. Applegate

MS Student, Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign

This report was submitted to the Collection Development Committee of the University of Illinois at Urbana-Champaign Library, 1/28/2016. It analyzes data on the online usage of the library’s digital collections from July 30, 2014 to July 30, 2015, and is intended to inform discussions within the library of digital collection development and access strategies.

Contents

Web Analytics in Libraries.....	2
UIUC Local Digital Collections	3
Digitized Books.....	4
Digitized Newspapers	6
Digitized Special Collections.....	7
Digital Scholarship.....	8
Finding Collections	10
Next Steps	11
References	11

URI: <http://hdl.handle.net/2142/89704>

Web Analytics in Libraries

Much has been written, particularly in the past decade, on the need for consistent and effective digital collection assessment in libraries. As recently as 2015, Duncan and O’Gara stated, “Although the forms of scholarship are changing, the framework for evaluating this new landscape of scholarly outputs is in its infancy.” Discussions of such evolving evaluative frameworks frequently make reference to web metrics and analytics. For the purpose of this report, **web metrics** is loosely defined as the general study of online behavior to answer questions about web users, whereas **web analytics** refers to the application of web metrics to usage data culled from a specific website, webpage, or group of pages. Web analytics data indicate things like how often users visit a particular site, what devices they use when accessing the web, and where they are located geographically—all data of interest to librarians wishing to assess the impact of collections they make available online. According to E. Kelly, “The second most prevalent topic in DL (digital library) assessment literature concerns the use of Web data to analyze usage and search patterns” (2014), a sentiment shared by Prom, who wrote in his discussion of web metrics for archival resources, “We can use analytics tools to improve our websites, to make effective decisions regarding online services, and to improve our users' experiences” (2011).

Researchers have however pointed out limits to the usefulness of web analytics data in digital collection usage analysis. Critically, web analytics numbers say nothing about user intention or motivation when accessing digital collections (Covey, 2002). For example, if both paper and digital copies of something exist, web analytics data do not explain why users select one over the other (Fleming-May & Grogg, 2010). As raw data, web metrics are rarely unambiguous (Khoo, et al., 2008), because analytics systems measure only the quantity of clicks on web pages rather than the quality of the user experience (Wacha & Wisner, 2011), not to mention that incomplete or inconsistent configuration of analytics software often produces incomplete data (Kyrillidou, et al., 2014). In other words, analytics data are not easily interpreted (Moore, et al., 2012).

These shortcomings notwithstanding, web analytics data do provide librarians with valuable information about the performance of their collections online. A thoughtful analysis of gathered analytics data can lead to informed decisions about website design, re-design, and alterations to enhance the user experience and to increase the discoverability of collections (Yang & Perrin, 2014). As demonstrated in a case study at the Ohio State University Libraries, web analytics data can be crucial in an assessment of usage and usability (Black, 2009). Another project conducted at the Kraemer Family Library at the University of Colorado, Colorado Springs made use of click analytics tools to identify and recommend links for removal (Farney, 2011) from a website. Finally, clear web analytics usage data are often relevant to funders and other stakeholders in the success of library digital collections (Eccles, et al., 2012).

This research suggests that librarians who wish to conduct web analytics studies need to do so with full awareness of the problems inherent to web statistics. As of 2015, best practices in this area are still emerging. Khoo and Donahue recommend normalization of disparate web analytics data in order to interpret meaningful trends (2007). Importantly, the Digital Library Federation presented "Collaborative Efforts to Develop Best Practices in Assessment: A Progress Report" at their 2015 Forum, detailing a multi-institutional effort to define best practices for accurately gauging “return on investment” using analytics data.

This report builds upon the findings cited above, and utilizes emerging best practices in its analysis of UIUC digital library collections.

UIUC Local Digital Collections

This report presents administrative data (number of collections, total items) and web analytics usage data (sessions, users, page views) from July 30, 2014 to July 30, 2015 for the following sites that provide access to locally managed digital collections:

- Digitized special collections materials in the **Digital Image Collections** (<http://imagesearchnew.library.illinois.edu/>)
- Digitized books and serials in the **HathiTrust Digital Library** (<https://www.hathitrust.org/>)
- Digitized newspapers in the **Illinois Digital Newspaper Collections** (<http://idnc.library.illinois.edu/>)
- Digital scholarship in the **IDEALS** institutional repository (<https://www.ideals.illinois.edu/>)

The library contributes digital collections to more digital library sites than these; however, as the sections below shall describe, these external services do not all provide reliable web analytics data.

Digital Library Site	Content Type	Number of Collections	Total items	Sessions	Users	Page Views
Digital Image Collections	manuscripts, documents, photographs, maps	24 collections	10,000 “objects” (45,941 images)	22,742	19,108	66,625
Illinois Digital Newspaper Collections	newspapers	39 newspaper titles	1,119,638 page images	393,715	322,905	1,645,314
IDEALS	research papers and other UIUC content	60 institutional “communities”	125,414 items	463,318	376,523	1,267,431
HathiTrust Digital Library	monographs and serials	3 UIUC digitization “streams”	408,911 UIUC books (among over 13,000,000)	Not available for UIUC-only	Not available for UIUC-only	3,968,905 for UIUC content (173,688,564 for entire repository)

Table 1. General statistics across several UIUC digital library services, July 30, 2014-July 30, 2015.

With differences from system to system in the way collections and the content they serve are structured, it is difficult to draw fair baseline comparisons of impact between them. This ambiguity is a central challenge to interpreting analytics data.

Digitized Books

Since 2007, the library has digitized over 383,510 volumes with corporate partners such as Google, 57,685 volumes in partnership with the Internet Archive, and several thousand volumes locally. The library's goal is to deposit all public domain books it scans into the HathiTrust Digital Library, and has adopted workflows to this effect.

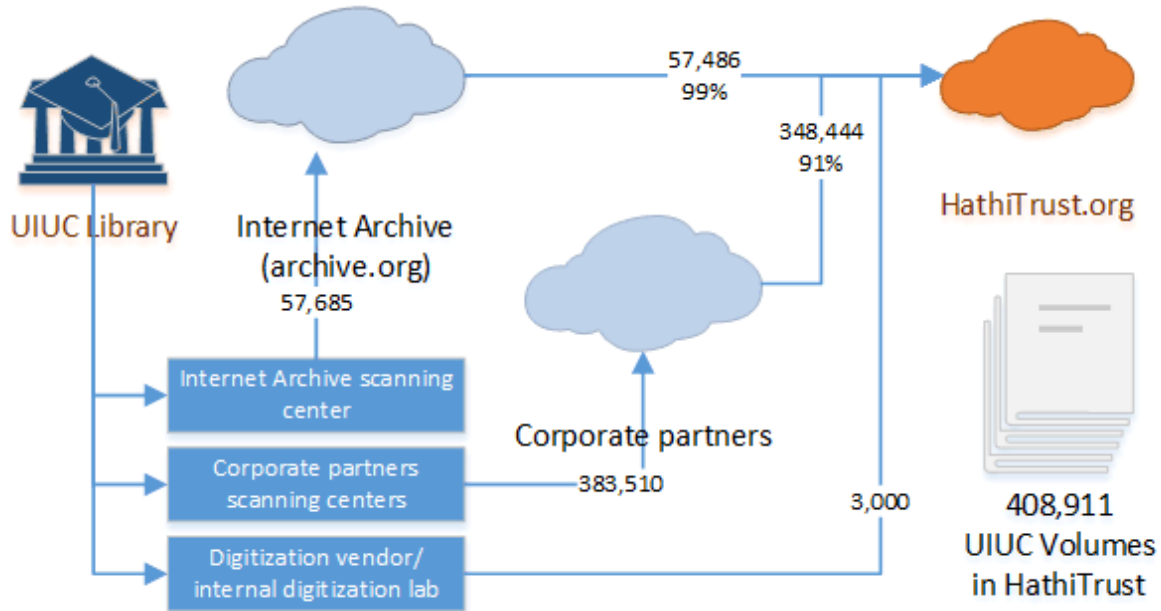


Figure 1. The path books take from library to digitization to availability on the web.

The library has developed a **Book Tracker** (https://medusa.library.illinois.edu/book_tracker/items) to show which digitized volumes have made it into the **Internet Archive** and the **HathiTrust Digital Library**. The **Internet Archive** has a page for download statistics for UIUC volumes (https://archive.org/details/university_of_illinois_urbana-champaign). The current top ten from the library's contributions are:

downloads	Title
40K	Devises et emblems anciennes & modernes, tirees de plus celebres auteurs
24K	Names of foreigners who took the oath of allegiance to the province and state of Pennsylvania, 1727-1775, with the foreign arrivals, 1786-1808
17K	Vol 1946-1951: Commencement program (UIUC)
16K	Vol 2: History of the Swedes of Illinois
16K	Vol 1978-1979: Faculty publications and doctoral dissertations (UIUC)
15K	History of the early settlers of Sangamon County, Illinois
14K	Vol 2: Original Sanskrit texts on the origin and history of the people of India, their religion and institutions
14K	Vol 2: Bibliografia polska XIX. stulecia; lata 1881-1900
14K	Pronunciation guide for Illinois place names
13K	Vol 5: Bibliografia polska. [Stólecie XV-XVIII]

Table 2. Top downloads of UIUC items in the Internet Archive.

Digitized Newspapers

The Illinois Digital Newspaper Collections (IDNC, <http://idnc.library.illinois.edu/>) consist of digital facsimiles of historical newspapers and trade journals, primarily from Illinois, dating from 1865-1922. Newspapers include those digitized locally and with support of the National Digital Newspaper Program (NDNP, <http://www.loc.gov/ndnp/>).¹ The IDNC is powered by a local instance of the Veridian (<http://www.veridiansoftware.com/>) digital newspaper software platform. Google Analytics statistics are available for each IDNC newspaper title. Titles are grouped on the IDNC site by thematic areas. The **Illinois Newspaper Collection** (10 titles) contains digital facsimiles of historic Illinois newspapers. The **Farm, Field and Fireside Collection** (20 titles) contains digital facsimiles of historic agricultural newspapers published in the United States. The **American Popular Entertainment collection** (3 titles) contains digital facsimiles of historic newspapers and trade journals published for the entertainment industry in the US between 1853 and 1929. The **Collegiate Chronicle** (6 titles) is a prototype student newspaper repository aggregating newspapers from colleges and universities across the USA.

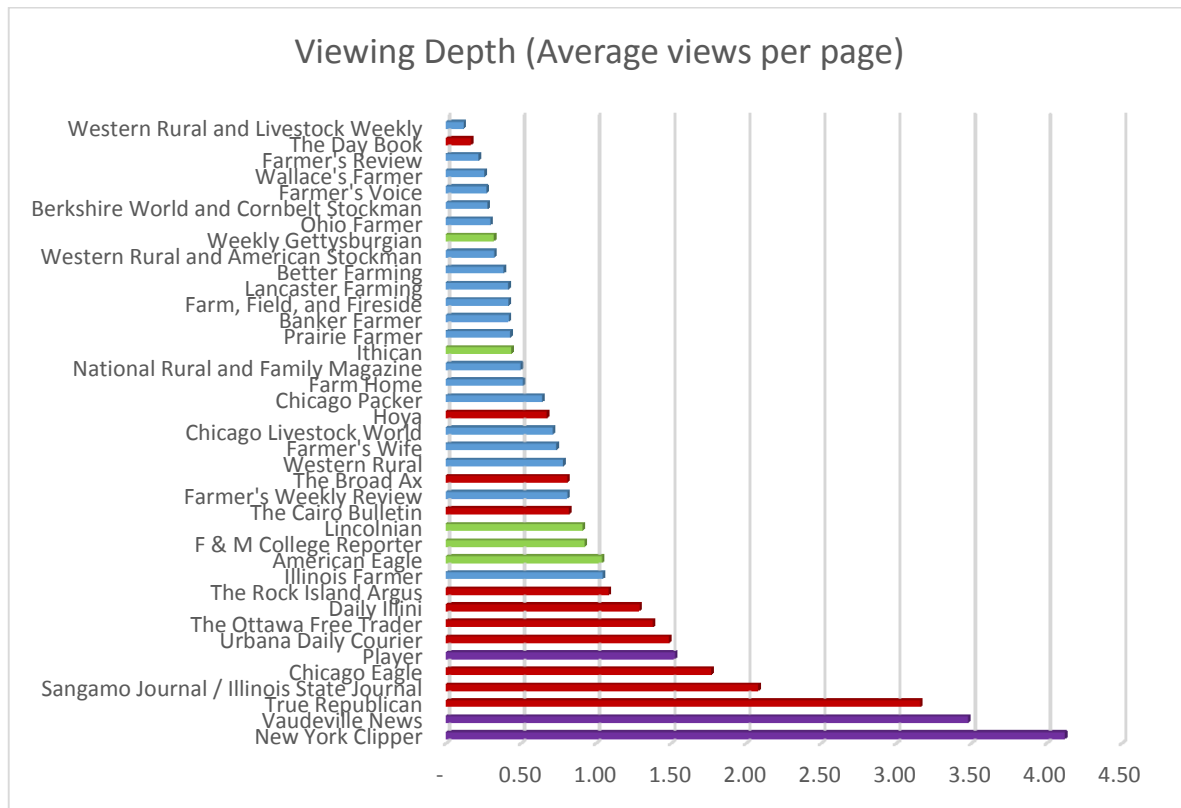


Figure 5. The authors calculated "viewing depth" by dividing the number of pages in each title into the total number of page views for that title during the analysis period. This figure indicates degrees of public interest in these titles.

Web usage numbers show high levels of interest in the library's **vaudeville newspapers** and its **Illinois town papers**, and lower interest in **college newspapers** and **farm newspapers**.

¹ UIUC also deposits its NDNP-digitized titles in Library of Congress's *Chronicling America* (<http://chroniclingamerica.loc.gov/>) portal. *Chronicling America* does not provide useful web analytics statistics to its partners. The statistics in this report are for the IDNC site only.

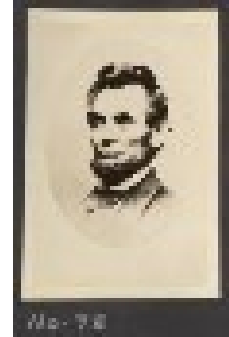
Digitized Special Collections

UIUC makes digitized special collections visual materials available in a Digital Image Collections (<http://imagesearchnew.library.illinois.edu/>) site driven by a locally hosted instance of OCLC's ContentDM (<https://www.oclc.org/contentdm.en.html>) platform. These collections feature historical photographs, maps, prints, letters, and other archival items. Among its collections, the following evinced clear levels of interest:

The Faces of Abraham Lincoln

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/lincoln>

108 items, 3,891 total page views



University Honors--The Bronze Tablets

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/bronze>

91 items, 2,952 total page views

French World War I Posters

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/wwposters>

105 items, 1,784 page views



Teaching with Digital Content (Digital Cultural Heritage Community)

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/tdc>

2,078 items, 2,798 page views

James Edward Myers World War I Sheet Music Collection

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/myers>

6,044 items, 4,551 page views

Motley Collection of Theatre and Costume Design

<http://imagesearchnew.library.illinois.edu/cdm/landingpage/collection/motley>

4,788 items, 3,142 page views

In total, the Digital Image Collections comprise more than 41,000 page images, and received about 18,000 page views for the year.

Digital Scholarship

Since 2007, the Illinois Digital Environment for Access to Learning and Scholarship (IDEALS, <https://www.ideals.illinois.edu/>) has served as UIUC's institutional repository. Powered by the open source DSpace software platform, IDEALS provides access to research papers, theses and dissertations, and other items deposited by the campus community, as well as to gray literature digitized by the library. Approximately 40,000 of IDEALS' 125,000 items were downloaded in the course of the year. The table below shows the most popular of these items. While a certain number of downloads are triggered by web crawler traffic, the median number of downloads per item for this period was 27 and the mean was 87. The numbers below, well above these levels, indicate online patron interest beyond any numeric bloat caused by "bots."

Table 3. Top 15 most frequently downloaded items in IDEALS, July 30, 2014-July 30, 2015 (items of "gray literature" distinguished by gray cells).

Title	downloads	work type
Successful resource sharing in academic and research libraries in Illinois: lessons for developing countries (9 files) ²	73,777	thesis
Estrategias para Motivar el Aprendizaje Colaborativo en Cursos a Distancia	21,444	unpublished conference paper
The product life cycle : its role in marketing strategy/some evolving observations about the life cycle	19,452	working paper (digitized by Internet Archive; however, the IA version has been downloaded only 870 times)
El aula virtual: Usos y elementos que la componen	18,574	unpublished conference paper
Heat Transfer in Refrigerator Condensers and Evaporators	16,279	technical report
Gender Differences in Leadership	16,117	published paper (in a journal with backfiles in IDEALS)
Leadership and employee motivation	12,229	published conference paper
Managing Resistance to Change	12,083	published paper (in a journal with backfiles in IDEALS)
From love at first sight to soul mate: Romantic ideals in popular films and their association with young people's beliefs about relationships	10,928	dissertation
Ethics and Professional Responsibility in Computing	9,505	published book chapter
Research on teaching reading comprehension (2 files)	9,014	technical report
UIUC Women's Crew: Origins, History, and Progress (4 files)	8,873	student project from campus Ethnography of the University Initiative
Analyses of differences between written and oral language (2 files)	8,841	technical report
Smartphones in the U.S.: Market Analysis	8,553	unpublished report
Riveted And Bolted Structural Joints	8,490	technical report

² Statistics for items with multiple files may receive some inflation compared to one-file items, as download numbers account for all downloads of files associated with a single item.

IDEALS statistics tell an interesting story about how people discover UIUC digital scholarship. The “Heat Evaporators and Condensators” report from the table above was downloaded 16,279 times over the course of the analysis year. This figure was provided by IDEALS’ underlying DSpace software³. Google Analytics however shows that the landing page for this same item only received 115 page views during the analysis period. This is because people are discovering the item via metadata exposed to external services like Google and Google Scholar, and are downloading the PDF document without ever seeing the IDEALS interface. In other words, the overwhelming majority of people who discover IDEALS content never actually view the IDEALS website.

Table 4. Top 10 communities in IDEALS, based on average number of downloads per item (for communities with 30 or more items).

Community Name	downloads	items	average downloads per item
Illinois Research and Scholarship (Open Community)	125,788	408	308
Illinois Fire Service Institute (IFSI)	12,973	57	228
IDEALS Initiative	6,094	32	190
Graduate School of Library and Information Science	992,521	6,305	157
Arms Control, Disarmament, and International Security (ACDIS)	10,492	72	146
Literary Gulag - Diana E. Sheets	6,218	45	138
European Union Center (EUC)	4,009	30	134
iSchools	125,795	967	130
University of Illinois at Springfield	25,709	198	130
Ethnography of the University Initiative	89,913	694	130

The top communities in IDEALS, as well as the top items, show that online patrons are most interested in IDEALS content that is unavailable elsewhere, with an emphasis on gray literature not available through traditional publishing channels.

³ For access to statistics on IDEALS items, simply append “/report” to the handle for any item: <https://www.ideals.illinois.edu/handle/2142/9750/report>

Finding Collections

How do users find their way to UIUC collections sites? Google Analytics distinguishes between **organic** traffic from web searches, **referral** traffic from external links, **direct** traffic from URLs typed into browsers or linked from documents, and **social** traffic from social media platforms.

Figure 6 (right) shows that many patrons discover content through Google or other web browser searches. This is

particularly true for **IDNC** newspaper (77%) and **IDEALS** institutional repository (69%) items. These organic web search results are lower for **HathiTrust**, which received 43% of its traffic from the web, and 41% from referrals, largely from library catalogs. **Social media** referrals, however, remain consistently below 2% across all four platforms and do not constitute a major driver of traffic to UIUC digital collections.

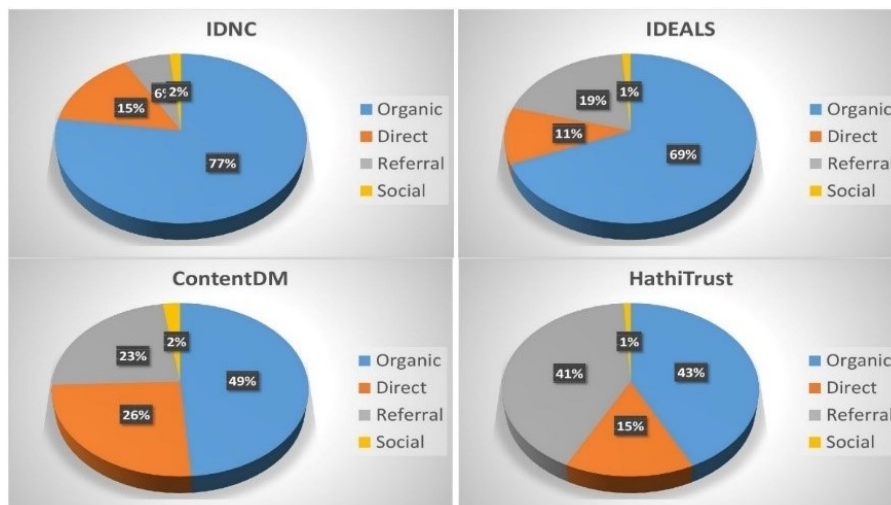


Figure 6. Web traffic. (NOTE: HathiTrust numbers are for entire HT, not just UIUC items).

Table 5. Top Ten Referrers per digital library site (excluding “organic” web searches)

Digital Image Collections (6,361 sessions)	IDNC (31,409 Sessions)	IDEALS (95,970 Sessions)	HathiTrust (4,475,067 Sessions)
images.library.uiuc.edu (31%)	en.wikipedia.org (11%)	scholar.google.com (21%)	onlinebooks.library.upenn.edu (18%)
dp.la (24%)	familyhistorydaily.com (8%)	library.illinois.edu (5%)	babel.hathitrust.org (13%)
t.co (Twitter) (2%)	freenewspaperarchives.us (7%)	search.grainger.illinois.edu (3%)	worldcat.org (6%)
vm136.lib.berkeley.edu (2%)	joinerhistoryroom.org (6%)	oatd.org (2%)	hathitrust.org (3%)
cluster3.lib.berkeley.edu (2%)	m.facebook.com (5%)	scholar.google.co.uk (2%)	clio.columbia.edu (1%)
m.facebook.com (1%)	researchguides.net (5%)	cee.illinois.edu (2%)	en.wikipedia.org (1%)
facebook.com (1%)	l.facebook.com (5%)	illinois.edu (2%)	dp.la (1%)
l.facebook.com (1%)	elephind.com (4%)	schools.org (1%)	mirlyn.lib.umich.edu (1%)
pinterest.com (1%)	facebook.com (3%)	scholar.google.co.in (1%)	psu.summon.serialssolutions.com (1%)
reddit.com (1%)	dekalb.illinoisgenweb.org (3%)	search.grainger.uiuc.edu (1%)	oskicat.berkeley.edu (1%)

Table 5 (above) shows top referral sources *other than* search engines—these numbers comprise the **direct**, **referral**, and **social** categories of Figure 6. They show that exposure of item metadata in the **Digital Public Library of America** (dp.la) has driven 24% of referral traffic back to the **Digital Image Collections**. In addition, **IDNC** staff efforts to produce informative **Wikipedia** articles linked to digitized newspapers have driven 11% of all referral traffic to the site. Statistics also demonstrate that **IDEALS** staff efforts to expose metadata to Google Scholar indexing have been successful in driving a significant number of online patrons to IDEALS items.

Next Steps

The authors will share this report with the library's Collection Development Committee (CDC), and will deposit it in IDEALS for broader access. If the CDC requests future iterations of this report, the authors recommend the following steps:

- Normalize settings on all library Google Analytics accounts to ensure optimal data collection.
- Synchronize data collection period with fiscal year for consistency with other assessment activities.

Please address comments to Kyle Rimkus (rimkus@illinois.edu).

References

- Bass, Kristin M., Cassidy Puckett, and Saul Rockman. "Models of Digital Collection Use in a University Community." *Educational Technology* 48.1 (2008): 44. Print.
- Black, Elizabeth L. "Web Analytics: A Picture of the Academic Library Web Site User." *Journal of Web Librarianship* 3.1 (2009): 3–14. *CrossRef*. Web.
- Cheri Jeanette Duncan, and Genya Morgan O'Gara. "Building Holistic and Agile Collection Development and Assessment." *Performance Measurement and Metrics* 16.1 (2015): 62–85. *emeraldinsight.com (Atypon)*. Web.
- Covey, Denise Troll. *Usage and Usability Assessment: Library Practices and Concerns*. Council on Library and Information Resources, 2002. Web. 9 July 2015.
- Fleming-May, Rachel A., and Jill E. Grogg. "Chapter 1: Assessing Use and Usage." *Library Technology Reports* 46.6 (2010): 5–10. Print.
- Kathryn E. Eccles, Mike Thelwall, and Eric T. Meyer. "Measuring the Web Impact of Digitised Scholarly Resources." *Journal of Documentation* 68.4 (2012): 512–526. *emeraldinsight.com (Atypon)*. Web.
- Kelly, Elizabeth Joan. "Assessment of Digitized Library and Archives Materials: A Literature Review." *Journal of Web Librarianship* 8.4 (2014): 384–403. *Taylor and Francis+NEJM*. Web.
- Khoo, Michael et al. "Using Web Metrics to Analyze Digital Libraries." *In Proceedings of the Joint Conference on Digital Libraries* (2008): 375–384. Print.
- Khoo, Michael, and Robert A. Donahue. "Evaluating Digital Libraries with Webmetrics." ACM Press, 2007. 484. *CrossRef*. Web. 8 July 2015.
- Kyrillidou, Martha et al. "Measuring Up: Assessing Accuracy of Reported Use and Impact of Digital Repositories." (2014): 1–10. Print.
- Moore, Will, Nancy MacCreery, and Martin Marlow. *Usage Measurements for Digital Content*. N.p., 2012. *Google Scholar*. Web. 9 July 2015.
- Prom, Christopher J. "Using Web Analytics to Improve Online Access to Archival Resources." *The American Archivist* 74.1 (2011): 158–184. Print.
- Wacha, Megan1 megan.wacha@gmail.com, and Meredith1 Wisner. "Measuring Value in Open Access Repositories." *Serials Librarian* 61.3-4 (2011): 377–388. *EBSCOhost*. Web.
- Yang, Le, and Joy M. Perrin. "Tutorials on Google Analytics: How to Craft a Web Analytics Report for a Library Web Site." *Journal of Web Librarianship* 8.4 (2014): 404–417. *EBSCOhost*. Web.