

Cornell University
ILR School

Practical Technology for Archives

Volume 1 | Issue 1

Article 2

11-2013

Using Google Analytics Data to Expand Discovery and Use of Digital Archival Content

Michael Szajewski
Ball State University

Follow this and additional works at: <https://digitalcommons.ilr.cornell.edu/pta>



Part of the [Archival Science Commons](#)

Thank you for downloading an article from DigitalCommons@ILR.

[Support this valuable resource today!](#)

This Article is brought to you for free and open access by DigitalCommons@ILR. It has been accepted for inclusion in Practical Technology for Archives by an authorized administrator of DigitalCommons@ILR. For more information, please contact catherwood-dig@cornell.edu.

If you have a disability and are having trouble accessing information on this website or need materials in an alternate format, contact web-accessibility@cornell.edu for assistance.

Using Google Analytics Data to Expand Discovery and Use of Digital Archival Content

Description

This article presents opportunities for the use of Google Analytics, a popular and freely available web analytics tool, to inform decision making for digital archivists managing online digital archives content. Emphasis is placed on the analysis of Google Analytics data to increase the visibility and discoverability of content. The article describes the use of Google Analytics to support fruitful digital outreach programs, to guide metadata creation for enhancing access, and to measure user demand to aid selection for digitization. Valuable reports, features, and tools in Google Analytics are identified and the use of these tools to gather meaningful data is explained.

Keywords

digital archives, Google Analytics, digital outreach, data

Using Google Analytics Data to Expand Discovery and Use of Digital Archival Content

Michael Szajewski

Ball State University

This article presents opportunities for the use of Google Analytics, a popular and freely available web analytics tool, to inform decision making for digital archivists managing online digital archives content. Emphasis is placed on the analysis of Google Analytics data to increase the visibility and discoverability of content. The article describes the use of Google Analytics to support fruitful digital outreach programs, to guide metadata creation for enhancing access, and to measure user demand to aid selection for digitization. Valuable reports, features, and tools in Google Analytics are identified and the use of these tools to gather meaningful data is explained.

Introduction

Recent scholarship focused on the intersection of libraries and web analytics has highlighted the value of web analytics tools in the enhancement of online library services and content. Web analytics has been traditionally been defined as “the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage.”ⁱ By using web analytics to understand web users’ interaction with library web content, information professionals can make more informed decisions regarding the development of digital library services and platforms. Web use data can inform site design, selection of content, and promotion and outreach strategies; it can also justify digital library expenditures by demonstrating evidence of use of web-based services and content. Given the value of web analytics for information service providers, information professionals must encourage a more widespread

practice of web data gathering and analysis amongst the profession to ensure sustained success of digital library initiatives.

Web analytics provide unique opportunities for archivists developing and maintaining freely available digital archives and repositories. These online, open-access collections differ from traditional proprietary resources that are limited to a specific library user community. As Loftus discusses, the use of page views as an indicator of success is problematic for traditional library institutions as their potential user base is more limitedⁱⁱ. Freely available digital repositories, however, seek to connect all potentially-interested web searchers with relevant archival assets. As a result of this critical difference in target audience, managers of library websites providing access to proprietary resources to a defined community and managers of freely available digital archives and repositories will likely approach the interpretation of web use data with different aims. Given the large potential audience and the high research impact ceiling for these freely available repositories, archivists have significant, unique opportunities to use web analytics to facilitate the use of and maximize the research value of their digital collections by driving traffic to their content.

This article will provide four practical examples of applied data analysis that uses Google Analytics, a freely available web analytics tool. These examples, pertaining specifically to freely accessible online archives and repositories, will describe several of the many ways that archivists can expand use and discovery of digital archival assets by web users. How can the effectiveness of digital outreach initiatives be evaluated and optimized? How can archivists uncover new digital outreach opportunities? How can the effectiveness of descriptive metadata in facilitating asset discovery be improved? How can digital archivists make more informed selection decisions in digital collection building? Each of these four topics will be addressed through the definition of relevant Google Analytics metrics, an explanation of methods for extracting and analyzing data gathered using the Google Analytics interface, and proposed strategies to use this data analysis to increase the use and visibility of a digital archival repository's materials. Each suggested strategy will be supported with brief examples from the author's analysis of

web use data for the Ball State University Digital Media Repository (DMR), a freely available CONTENTdm-based digital archives developed by the Ball State University Librariesⁱⁱⁱ.

About Google Analytics

Released in 2006, Google Analytics is a freely available, user-friendly web analytics solution developed by Google that provides users with highly interactive, flexible, and detailed web analytics reporting^{iv}. Through Google Analytics' main web-based interface, the user seeking to gather and analyze data builds a query based on specific metrics, defined as "individual measurements of visitor activity on your site," and dimensions, which "break down metrics across some common criteria, such as country or browser."^v Google provides users with a wealth of robust yet accessible supporting documentation, including a detailed reference page outlining various metrics and dimensions that are available through the Google Analytics interface, demystifying web analytics language for beginning users^{vi}. The Google Analytics installation and set-up process is thoroughly and clearly described in Google's supporting documentation for the Analytics service^{vii}. The creation of a Google Analytics account provides the would-be Analytics user with a tracking code snippet to be placed in the web page or web page template's HTML document.

The service separates its standard reporting features into broad categories that group various reporting options by function. Audience reports allow a Google Analytics user to obtain data about the users of the web site being tracked, including their geographic location, the duration of their visits, the number of pages accessed in a visit, and the hardware and software they use to access the site. Acquisition reports allow Google Analytics users to obtain data regarding the ways in which users discover and access the site's content, including a list of web search engine searches that brought users to the site and data regarding sites that referred users to the site directly. Behavior reports provide data regarding user behavior once the use has arrived on the site, illuminating, for example, which pages on the site are most heavily viewed and

which pages on the site most frequently serve as a site's entry point^{viii}. In 2011, Google Analytics added a Realtime component to its interface which allowed for immediate access to data to support timely monitoring of site performance^{ix}.

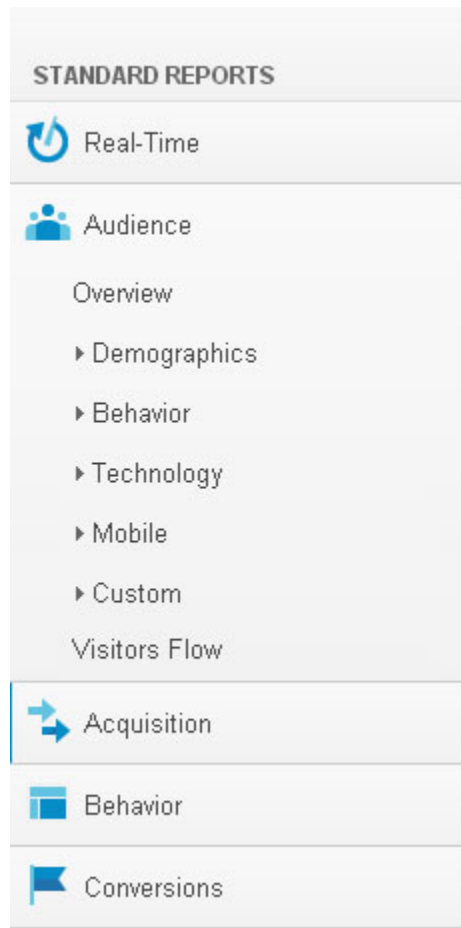


Figure 1: Google Analytics reports are organized into categories by function.

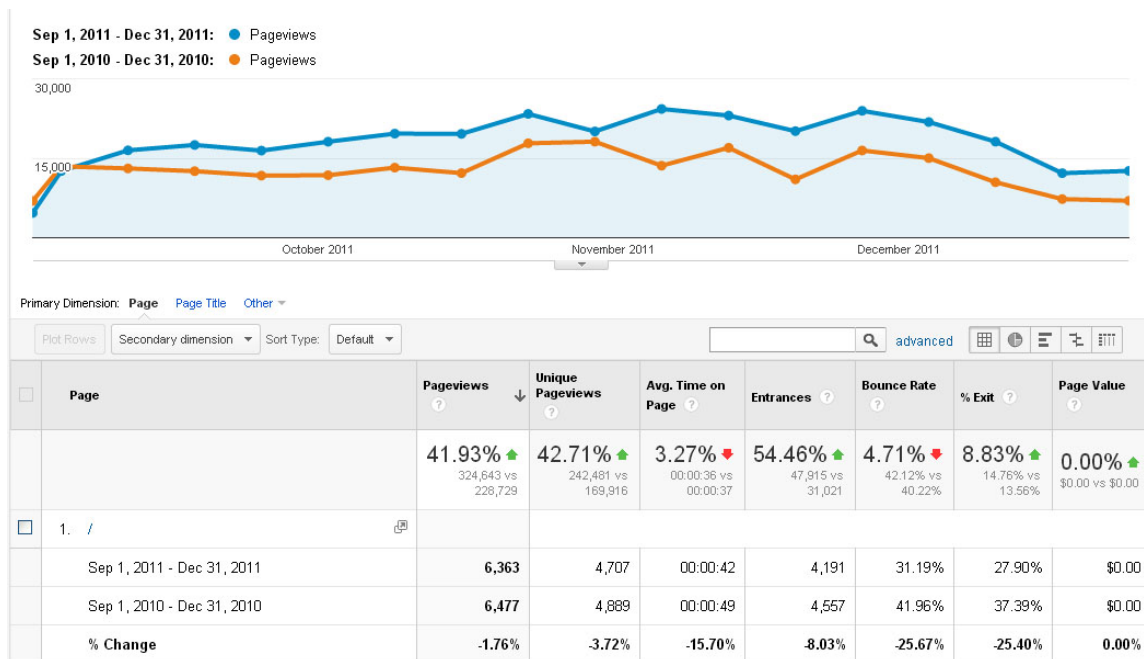


Figure 2: The Google Analytics interface combines data reporting with valuable, flexible visualizations.

Literature Review

A substantial body of information science literature documents the use of web analytics tracking services, including Google Analytics, to analyze the use of library web pages and digital libraries; some research in this area has been expanded into the field of archival studies as well. The literature in this field thoroughly documents the value of web data analysis and the process of developing a web analytics program at a library institution. Much of the published scholarship on the topic of web analytics in libraries has focused on the use of data to improve site usability and user experience. Sumner et al. documented the planning of a web metrics strategy for the National Science Digital Library, identifying goals for the initiative as well as potential challenges^x. Khoo et al. present a general overview of web metrics providers, valuable metrics, and factors affecting implementation, as well as brief implementation case studies involving four

digital libraries specifically focusing on session length data. Khoo writes that session length data is a problematic measure of visit quality in the context of libraries, as some visitors may be looking to recall one item quickly while others may pursue extensive research^{xi}. Fang et al. analyzes factors such as connection type and speed, browser type, screen resolution, page visits, and geographic statistics to better understand the Rutgers-Newark Law Library's user base and demonstrate that increasing the prominence of certain pages further promotes their use by digital library users^{xii}.

Black investigates the analysis of web log data at The Ohio State University Libraries' to provide basic information regarding web user interaction, technology, and content preferences^{xiii}. Paul investigates the implementation of Google Analytics on the University of Missouri's library webpages, reporting on the library's willingness to implement changes to links on pages but its reluctance to consider more widespread, large-scale changes based on data gathered^{xiv}. Loftus discusses the use of Google Analytics by the Health Sciences Libraries of the University of Minnesota, including the use of custom variable to track data from more complex sites. His analysis explains how the institution used web use data to inform decisions to feature underused pages more prominently and how subsequent follow-up data analysis showed that the redesign achieved its aims^{xv}. Hess summarizes the functionality and installation of Google Analytics, highlighting useful features of the services and more complex data extraction strategies, and reports on the use of Google Analytics to evaluate the effectiveness of the Illinois Harvest Portal discovery layer^{xvi}.

Betty describes the implementation of Google Analytics to study use of online library screencast tutorials at Regis University. Metrics regarding the user software and connection speed allowed for analysis of the accessibility of the screencast videos^{xvii}. Wagner and Arendt document the implementation of Google Analytics-based data analysis program to analyze use of the Morris Library at Southern Illinois University Carbondale website and inform a site redesign. They discuss challenges in interpreting Google Analytics data to draw meaningful conclusions about web use and to support design change^{xviii}. Turner emphasizes the importance of establishing performance goals

for a library website and recommends using Google Analytics goal conversion features to evaluate the successes of these goals^{xix}.

Members of the archival community have also produced scholarship on the usefulness of web analytics data in the managing of online archival content. Prom provides an in-depth study of the use of Google Analytics at the University of Illinois Archives to provide data regarding use of online finding aid content. Prom's study focuses on the use of data to inform site redesigns to improve user experience and encourage greater user engagement^{xx}. O'English describes the use of Google Analytics to study how library patrons used and discovered HTML archival finding aids at the Manuscripts, Archives, and Special Collections at the Washington State University Libraries, highlighting the importance of web traffic in finding aid visibility and discoverability^{xxi}. Ament-Gjevick studies the use of multiple tools for web data gathering, including Google Analytics and Quantcast, and demonstrates the value of using referral site metrics to understand web use and web use trends to enhance accessibility^{xxii}.

The Value of Discoverability and Visibility

Tracking the research amongst the information service community regarding the use of web analytics clearly shows that scholarship in this area is firmly rooted in studies conducted by librarians who study and manage traditional online library sites that provide a defined, limited community access to access to mainly-proprietary resources. These studies primarily focus on enhancing the experience of their community's online users by analyzing web use data to inform site redesign and improvement. Members of the archival community who manage or study freely available digital repositories and have explored the value of web analytics have demonstrated the influence of many of the aforementioned researchers through their emphasis on internal site performance and experience. In addition, they have also realized that, due to the uniqueness of their repositories, emphasis on online discoverability is also significant when studying analytics.

However, a review of research in this area demonstrated a significant lack of scholarship dedicated to the role of web analytics in developing a program specifically intended to increase the visibility, discovery, and use of digitized archival assets amongst a broad audience of web users. The literature has also vastly ignored the value of web analytics in informing the development of web outreach programs using social media and other Web 2.0 outlets. Efforts to maximize the visibility and use of their online materials are of great significance and value to the digital archivist. Visibility-raising efforts have the ability to increase the research impact and relevance of digital archives. However, digital archives are competing with a plethora of online materials and information for the interest and visitation of users who would be potentially interested in the subject matter documented. Thus, emphasis on discoverability is not only beneficial but essential in assuring the cultural relevance of digital archives.

The four questions and subsequent discussion below address four significant opportunities to increase the visibility of and traffic to digital archival objects that can be informed and evaluated through the analysis of data provided by Google Analytics. Each of the questions will be addressed through the description of specific queries in Google Analytics that can provide relevant data, the definition of metrics relevant to the query, and discussion of useful and productive analysis of the data returned from the queries. Discussion will also include relevant examples from the author's analysis of web use data for the Ball State University DMR.

How can the effectiveness of digital outreach initiatives be evaluated and optimized?

The exploration of social media sites and the development of Web 2.0 content to promote library materials and services have been thoroughly explored through both recent scholarship and practice in the library and archival communities. Through the use of Google Analytics, digital archivists can gather data documenting the amount of web traffic directed towards a digital asset or page via social media outlets and other

external sites and use this data to analyze the success of digital outreach initiatives and to improve the effectiveness of these efforts.

The Acquisition reporting section of Google Analytics provides access to valuable metrics describing the specific ways in which users discover a site's content. The All Referrals report documents traffic referred to a site via links on other websites, listing the referring source (the website from which the visitor came) and the number of visitors directed from that site over any period of time. Other valuable data regarding the length and depth of a visit, the percentage of new visitors attracted, and the bounce rate (the percentage of single-page visits^{xxiii}) is also presented for each individual source^{xxiv}. Each source can be individually selected and isolated, presenting a report that visualizes the number of visitors referred via that exact source for any specific, selected time and breaks down the traffic via that source into multiple, unique referral paths (the specific page on the referral site from which the traffic came^{xxv}), if applicable.

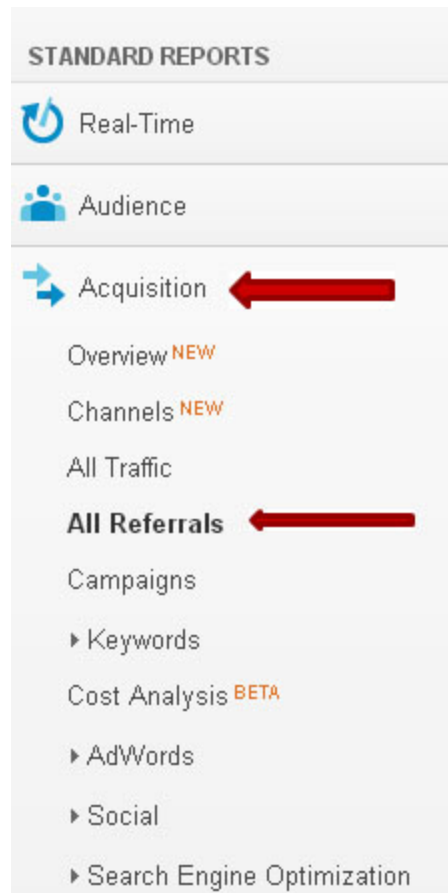


Figure 3: Accessing the Referrals report.

Plot Rows		Secondary dimension	Sort Type: Default	advanced		
<input type="checkbox"/>	Source	Visits ? ↓	Pages / Visit ?	Avg. Visit Duration ?	% New Visits ?	Bounce Rate ?
		58,266 % of Total: 32.66% (178,410)	8.66 Site Avg: 6.27 (38.15%)	00:04:16 Site Avg: 00:03:03 (39.70%)	73.41% Site Avg: 80.26% (-8.54%)	35.31% Site Avg: 50.83% (-30.53%)
<input type="checkbox"/>	1. en.wikipedia.org	18,902	6.11	00:02:44	84.85%	30.91%
<input type="checkbox"/>	2. cms.bsu.edu	13,771	15.53	00:08:24	49.12%	26.52%
<input type="checkbox"/>	3. artcyclopedia.com	5,675	3.55	00:01:06	84.93%	45.62%
<input type="checkbox"/>	4. facebook.com	2,825	9.17	00:04:37	65.06%	46.65%
<input type="checkbox"/>	5. liblink.bsu.edu	1,039	4.35	00:02:54	77.57%	59.87%
<input type="checkbox"/>	6. en.m.wikipedia.org	872	3.13	00:01:13	90.94%	30.85%
<input type="checkbox"/>	7. bsu.edu	827	6.72	00:03:17	67.96%	22.85%
<input type="checkbox"/>	8. eudocs.lib.byu.edu	750	3.62	00:02:30	89.47%	47.20%

Figure 4: Referrals report lists referring sources, number of visits per source, and other valuable data.

Data gathered from this report can be used to increase traffic to a site in several ways. The data can determine which referral sites are generally most fruitful in directing traffic to a digital library, allowing archivists to spend more resources promoting their materials via the most fruitful sources. The data can also show which specific instances of a particular type of outreach were most successful, allowing archivists to hypothesize about the factors that make a particular outreach instance attractive or successful. In promoting assets from the Ball State University DMR, Google Analytics data was used in both of these ways to inform digital outreach efforts to maximize traffic.

At Ball State University, the use of Wikipedia has been explored as a means of raising the visibility and discoverability of digital archival objects in the Digital Media Repository. The efforts were heightened in 2011 when 144 links to DMR objects from relevant, already-existing Wikipedia pages with links being typically added to the

External Links, References, and Further Reading sections of Wikipedia articles. Google Analytics was used to track the number of visitors who arrived at the DMR via Wikipedia. The data gathered from the All Referrals report demonstrated that, from the months of September to December, 2011, 5,663 visitors arrived to the DMR via Wikipedia, representing 11.82% of our total traffic. This data was very encouraging and justified the continued use of Wikipedia to raise the visibility of the DMR and its assets.

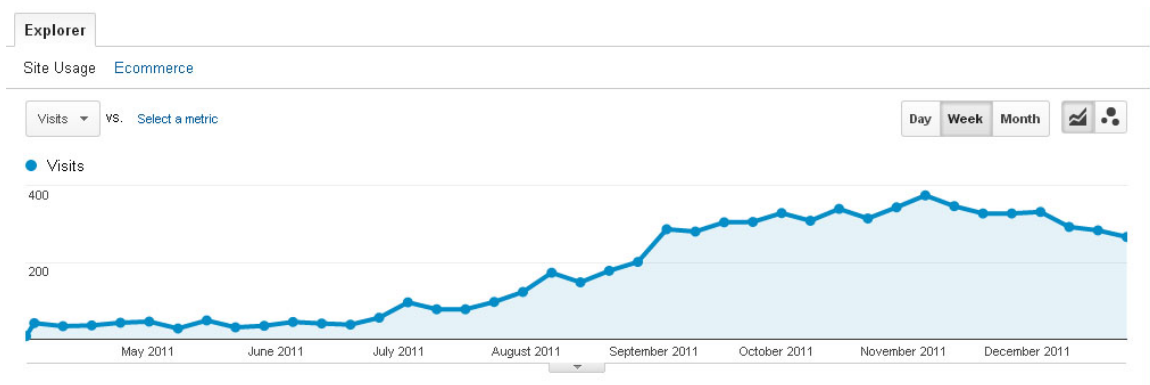


Figure 5: Data visualization shows growing number of referrals via Wikipedia.org as more links were added, supporting impact of Wikipedia outreach program on object visibility.

Data regarding the number of visitors who arrived via specific Wikipedia articles was available by selecting the general Wikipedia source in the All Referrals report. This data helped identify which types of Wikipedia articles attracted the most visitors to the DMR, showing that links to digitized historic sheet music and links to published, freely accessible works of scholarship tended to be especially high sources of traffic. Using this data, the adding of links to these two types of materials was further explored.

<input type="checkbox"/>	Referral Path	Visits [?] ↓	Pages / Visit [?]	Avg. Visit Duration [?]	% New Visits [?]	Bounce Rate [?]
		18,902 % of Total: 10.59% (178,410)	6.11 Site Avg: 6.27 (-2.49%)	00:02:44 Site Avg: 00:03:03 (-10.54%)	84.85% Site Avg: 80.26% (5.72%)	30.91% Site Avg: 50.83% (-39.19%)
<input type="checkbox"/>	1. /wiki/Let_Me_Call_You_Sweetheart ^[D]	1,277	6.69	00:02:36	64.37%	13.70%
<input type="checkbox"/>	2. /wiki/East_of_Eden_(novel) ^[D]	731	3.65	00:02:35	63.20%	48.97%
<input type="checkbox"/>	3. /wiki/William_Arms_Fisher ^[D]	721	8.89	00:04:05	87.38%	0.55%
<input type="checkbox"/>	4. /wiki/A_Perfect_Day_(song) ^[D]	683	8.76	00:03:47	87.99%	2.20%
<input type="checkbox"/>	5. /wiki/Nearer_My_God_to_Thee ^[D]	518	6.84	00:02:50	90.35%	1.93%
<input type="checkbox"/>	6. /wiki/Belle_Époque ^[D]	448	2.24	00:01:55	68.53%	47.10%
<input type="checkbox"/>	7. /wiki/Till_We_Meet_Again_(1918_song) ^[D]	410	7.78	00:03:28	90.73%	2.44%
<input type="checkbox"/>	8. /wiki/Sons_and_Lovers ^[D]	396	1.23	00:00:30	94.44%	88.38%
<input type="checkbox"/>	9. /wiki/The_Red_Pony ^[D]	382	4.84	00:02:48	93.46%	37.17%
<input type="checkbox"/>	10. /wiki/It's_a_Long_Way_to_Tipperary ^[D]	358	8.81	00:04:21	86.59%	1.96%

Figure 6: By selecting a specific source, data regarding all unique referral paths from that source is presented. Shown above are the top ten referral paths from within Wikipedia.org over a selected time.

How can archivists uncover new digital outreach opportunities?

Online outreach efforts initiated by librarians and archivists are only one component of the sum total of digital promotion that can raise the visibility of digital archival materials. Third party sharing, in which online users discover archival materials and link to and discuss them on any number of Web 2.0 or social media platforms, can be crucial in reaching new audiences for online archives. The web contains a plethora of communities and sites that would provide unique, creative targeted digital outreach initiatives, many of which are unexplored. By using the All Referrals report in Google Analytics, digital archivists can have access to data documenting third party sharing and its impact on web traffic. By analyzing the list of sources in this report, archivists can identify sites and online communities that were especially fruitful traffic sources; this

information can be used to develop archives-generated outreach efforts that can further raise the visibility of digital archival content.

At Ball State University, using Google Analytics to document third-party sharing shed light on new digital outreach opportunities for DMR assets. In the spring of 2012, the DMR published a collection of over 800 digitized historic photographs taken by photographer and editor Ralph Satterlee documenting the Indianapolis 500 from 1960 to 1974^{xxvi}. Within a year, links to materials from this collection were posted to four separate auto racing message boards, mb2501.proboards.com, indycarmodeling.proboards.com, trackforum.com, and forums.autosport.com/, by a third-party user. In total, links posted on these four traffic sources accounted for the discovery of DMR assets from this collection by 370 users from May 1, 2012, to May 1, 2013.

<input type="checkbox"/>	29. ddarchive.blogspot.com	147	28.50	00:10:43	53.74%	
<input type="checkbox"/>	30. mb2501.proboards.com 	142	16.35	00:06:33	44.37%	
<input type="checkbox"/>	31. iolug.org	134	4.72	00:02:22	44.03%	
<input type="checkbox"/>	32. csnbbs.com	133	2.65	00:01:29	91.73%	
<input type="checkbox"/>	33. bsudmrnews.blogspot.com	127	13.67	00:06:58	55.12%	
<input type="checkbox"/>	34. trackforum.com 	123	22.98	00:06:46	84.55%	
<input type="checkbox"/>	35. search.incredibar.com	117	2.06	00:00:47	93.16%	
<input type="checkbox"/>	36. forum.axishistory.com	116	12.57	00:05:11	88.79%	
<input type="checkbox"/>	37. irongarmx.net	113	11.56	00:02:32	88.50%	
<input type="checkbox"/>	38. cityofmuncie.com	112	1.82	00:00:35	94.64%	

Figure 7: Scanning the list of sources in the All Referrals report can uncover third-party sharing activities.

The discovery of the fruitfulness of these third-party activities provided evidence of the value of sports message boards as a venue for digital archives promotion. In the fall of 2012 and winter of 2013, the author posted links to bsufans.com, a Ball State University sports message board, promoting digitized films of historic Ball State football and basketball games that had been made available in the DMR. During the same one-year period, the message board directed 184 users to the DMR.

How can the effectiveness of descriptive metadata in facilitating asset discovery be improved?

Descriptive standards have traditionally played a singular role in determining how digital archival assets are described and what descriptive information should be created to facilitate access. However, as Schaffner points out, digital librarians and archivists must consider user needs and behaviors when describing digital objects. “One of several core competencies that special collections metadata librarians must have is ‘a keen understanding of users’ needs and preferences,’” she writes. “This is especially important now that discovery happens in multiple environments. Librarians and archivists need to manage archival collections by provenance, but also must describe what is in the collections for their users^{xxvii}.” Google Analytics provides digital archivists with an opportunity to gain an understanding of user needs by offering data regarding search engine queries that connected users with relevant digital archives content. This data can inform Search Engine Optimization (SEO), efforts to raise the discoverability of assets via a search engine, allowing archivists to describe digital assets to facilitate increased use.

Web traffic data provided in case studies by both O’English^{xxviii} and Prom^{xxix} demonstrate the overwhelming significance of search engine traffic in the use of digitized archival content. By utilizing the Organic Keywords report in Google Analytics, digital archivists can see a list of web search engine searches that successfully brought web users to digital archives content. This report is presented in the Acquisitions reporting section. By accessing the content of searches that connected web searchers

with archival resources, archivists can better learn how to describe assets in ways that makes them discoverable to a greater number of potential users.

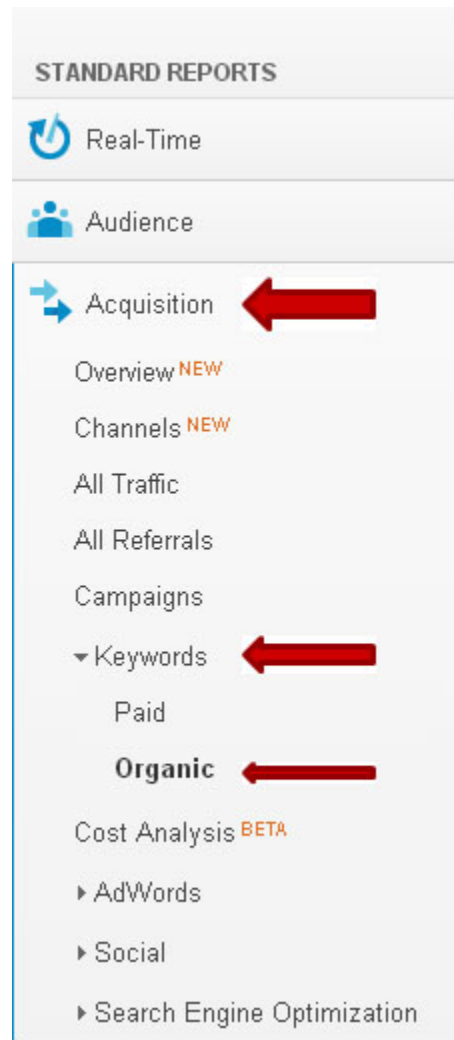


Figure 8: The Organic Keywords report provides access to keyword data.

<input type="checkbox"/>	Keyword	Visits [?] ↓	Pages / Visit [?]	Avg. Visit Duration [?]
		16,250 % of Total: 33.10% (49,091)	2.37 Site Avg: 3.17 (-25.32%)	00:01:26 Site Avg: 00:02:22 (-39.12%)
<input type="checkbox"/>	1. arm muscles	1,055	1.27	00:00:18
<input type="checkbox"/>	2. arm muscle anatomy	201	1.17	00:00:12
<input type="checkbox"/>	3. arm muscle diagram	180	1.14	00:00:07
<input type="checkbox"/>	4. arm muscles diagram	140	1.19	00:00:11
<input type="checkbox"/>	5. muscles of the arm	130	1.23	00:00:32
<input type="checkbox"/>	6. arm muscles anatomy	113	1.43	00:00:30
<input type="checkbox"/>	7. muscles in the arm	113	1.42	00:00:24
<input type="checkbox"/>	8. jody donovan	99	1.12	00:00:15
<input type="checkbox"/>	9. jody donovan singer	95	1.22	00:00:12
<input type="checkbox"/>	10. kidney diagram	95	1.22	00:00:20

Figure 9: Organic Keyword report provides list of and data regarding search engine queries that brought web users to a site.

At Ball State University, the Organic Keywords report was used to inform the research and creation of descriptive metadata for a large, growing collection of photographs documenting Ball State's history that were digitized and added to the DMR. As the collection grew and more assets were added to the DMR, this report was used to evaluate the effect of researching names of individual persons in the historic photographs and adding them to the object metadata on web search engine traffic to the digitized photographs. To isolate the keywords that specifically produced hits on the Ball State University Campus Photographs collection, a more complex process was required. In the Behavior reporting section, the Landing Pages report (organized under Site Content) was selected, which shows a list of all of the first pages in each user's session^{xxx}. Keyword was selected as a secondary dimension, showing the search engine query that led users to the specific landing pages when applicable. The filter was then used to isolate landing pages only from the Ball State University Campus Photographs

collection. From June 1, 2013 to September 1, 2013, keyword data was available for 388 site entries via the Campus Photographs collection through search engine traffic. Of these 388 visits, 137 (35.31 %) were the result of a search engine query that involved an individual person name. This data suggests that researching and providing names of individuals pictured in photographs, when possible, had a significant impact on the visibility of and traffic to digitized historical photographs in this collection.

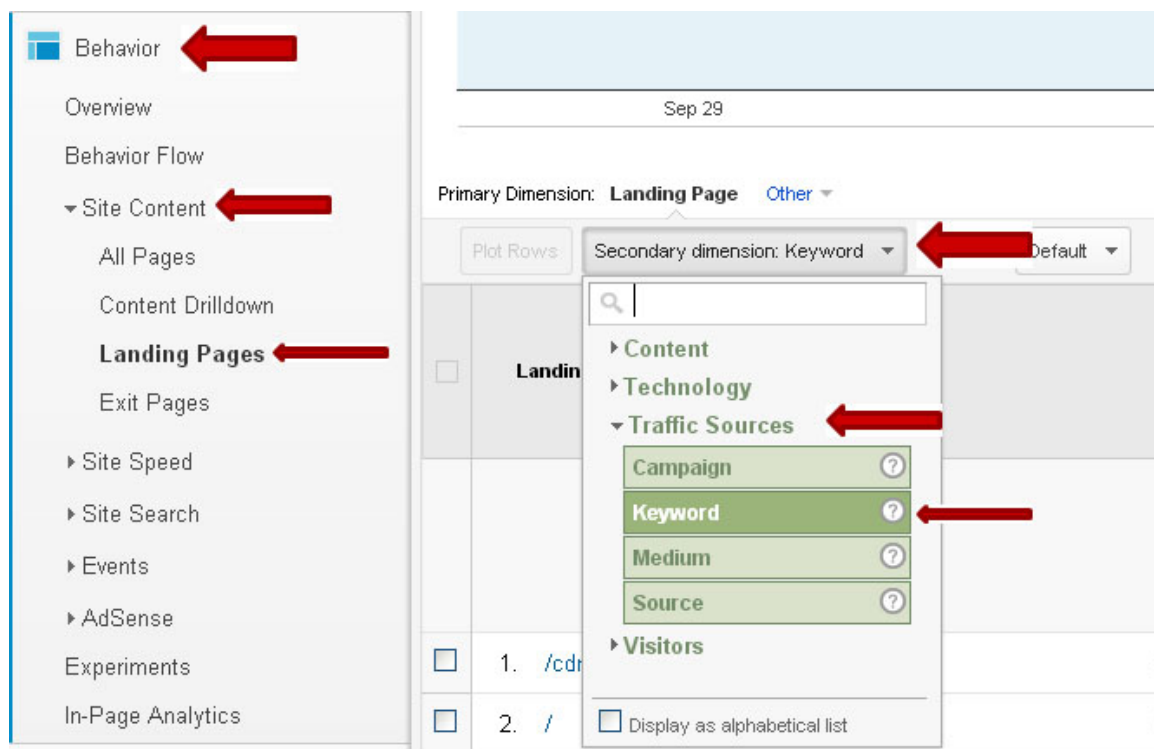


Figure 10: The Landing Pages report is selected with Keyword as a secondary dimension.

Plot Rows Secondary dimension: Keyword Sort Type: Default Advanced Filter ON X edit

Include Landing Page Containing collection/photo

and

+ Add a dimension or metric

Apply cancel

Figure 11: The Advanced Filter option isolates only landing pages from within the Campus Photographs collection.

How can digital archivists make more informed selection decisions in digital collection building?

As investment in digital archives initiatives continues, the selection of materials to make available online will remain a pivotal aspect of the management and growth of online archival repositories. By tracking and analyzing web analytics data, archivists can gain access to information to support informed decision making regarding selection for digitization. Gertz writes, on the matter of selection criteria, that “demand from users is vital. Digitizing and mounting materials publicly is a form of publishing, and success in publishing means knowing and targeting the audience^{xxxi}.” Archivists must not simply extrapolate digital demand from in-house use, as the user base for in-person patrons and online users varies greatly. Instead, they must seek to discover hard data to measure the demand for digitized archival assets.

Google Analytics’ Behavior reporting section offers reports that provide data regarding the specific pages that are most heavily visited on a website. The All Pages report (organized under Site Content) provides a list of all pages on a site that have been viewed, the number of page views that each page has received over any specified time

(with a page view being defined as any instance of the loading of a page by a user^{xxxii}), and other data regarding a user's interaction with that page^{xxxiii}. Pages can be viewed with the unique page URL or the page title as the primary dimension. The Landing Pages report provides similar data, but isolates only the pages through which a user first entered the site. This data can provide archivists with a thorough understanding of which items and collections are most heavily used and which are most successful in drawing users to the repository. The advanced filter option provides digital curators with the dexterity needed to isolate certain subsets of material based on common URL strings or page title elements.

Behavior

Overview

Behavior Flow

Site Content

All Pages

Content Drilldown

Landing Pages

Exit Pages

Site Speed

Site Search

Events

AdSense

Experiments

In-Page Analytics

Conversions

Sep 29

Oct 6

Primary Dimension: Page Page Title Other

Plot Rows

Secondary dimension

Sort Type: Default

<input type="checkbox"/>	Page
<input type="checkbox"/>	1. /cdm/ref/collection/AnatMod/id/211
<input type="checkbox"/>	2. /
<input type="checkbox"/>	3. /cdm/
<input type="checkbox"/>	4. /cdm/browseby?op=collection
<input type="checkbox"/>	5. /cdm/landingpage/collection/MnCnHSatFlm
<input type="checkbox"/>	6. /cdm/singleitem/collection/BSUArtists/id/2125/rec/10
<input type="checkbox"/>	7. /cdm/search
<input type="checkbox"/>	8. /cdm/search/collection/MusInst/searchterm/Musical instruments -- Latin America/field/all/mode/exact/conn/and/cosuppress/
<input type="checkbox"/>	9. /cdm/search/collection/MuseumBSU
<input type="checkbox"/>	10. /cdm/singleitem/collection/AnatMod/id/139

Figure 12: All Pages report showing pages listed by URL.

calendar year, of the 1,118,044 total page views logged on the DMR, 91,963 of them were recorded on Hague Sheet Music pages and objects. Assets from this collection were isolated in Google Analytics using the Advanced Filter feature. These archival sheet music objects were often discovered through Google and other search engines and were also easily integrated into our Wikipedia outreach strategy. The data gathered regarding these materials suggested the existence of a significant demand for historic sheet music amongst our web users, and as a result, digitization has begun on other sheet music collections in the Ball State University Archives & Special Collections to further meet this demonstrated demand.

All Traffic All Referrals Campaigns Keywords Paid Organic Cost Analysis BETA AdWords Social Search Engine Optimization Behavior Overview Behavior Flow Site Content All Pages Content Drilldown Landing Pages Exit Pages Site Speed Site Search Events AdSense	Primary Dimension: Page Page Title Other	
	Plot Rows	Secondary dimension Sort Type: Default
<input type="checkbox"/> Page		Pageviews ? ↓ 91,963 <small>% of Total: 8.23% (1,118,044)</small>
<input type="checkbox"/>	1. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=714&CISOBOX=0	1,940
<input type="checkbox"/>	2. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=920&CISOBOX=0	1,605
<input type="checkbox"/>	3. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=981&CISOBOX=0	1,501
<input type="checkbox"/>	4. /cdm4/document.php?CISOROOT=/ShtMus&CISOPTR=928&REC=11	1,445
<input type="checkbox"/>	5. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=1085&CISOBOX=0	1,295
<input type="checkbox"/>	6. /cdm4/document.php?CISOROOT=/ShtMus&CISOPTR=989&REC=12	1,093
<input type="checkbox"/>	7. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=982&CISOBOX=0	959
<input type="checkbox"/>	8. /cdm4/document.php?CISOROOT=/ShtMus&CISOPTR=720&REC=7	952
<input type="checkbox"/>	9. /cdm4/document.php?CISOROOT=/ShtMus&CISOPTR=1093&REC=15	949
<input type="checkbox"/>	10. /cdm4/doc_viewer.php?CISOROOT=/ShtMus&CISOPTR=649&CISOBOX=0	926

Figure 14: All Pages report isolating the Hague Sheet Music assets demonstrates the collection's strong performance.

Conclusion

Research has documented that Google Analytics can provide valuable data to enhance the understanding of a digital library user's experience and implement internal site improvements. However, the service also provides valuable data to inform the development and refinement of a digital outreach strategy to maximize the visibility of digital archives assets and to guide the description and selection of archival materials to promote widespread discovery and use of archival content. The features demonstrated provide archivists and digital library curators with easily-accessible, flexible data reports

that can enhance archivists understanding of the relationship between content and web users, both existing and potential. Exploration of web analytics should be encouraged as an activity in which archivists should engage as knowledge of users and content is elemental to decision-making, collection building, and outreach in archives. Through this practice, archivists can access the data needed to support decision-making necessary to maximize the research impact and relevance of digital archives.

About the Author

Michael Szajewski has served as Archivist for Digital Development and University Records at Ball State University since 2011. In this position, he develops and promotes digital collections, manages digitization projects, and oversees Ball State's university records collection area. He earned a Bachelor of Arts degree in history and communication studies from Northwestern University in 2009 and a Master of Library Science degree from Indiana University in 2010 with a concentration in archives and records management.

-
- ⁱ Web Analytics Association, Web Analytics Definitions. (2008), 3.
http://c.ymcdn.com/sites/www.digitalanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitions.pdf.
- ⁱⁱ Wayne Loftus, "Demonstrating Success: Web Analytics and Continuous Improvement," Journal of Web Librarianship 6, no. 1 (2012): 47, doi: 10.1080/19322909.2012.651416
- ⁱⁱⁱ Ball State University Digital Media Repository. Ball State University Libraries. accessed September 18, 2013, <http://www.libx.bsu.edu>
- ^{iv} "Our History in Depth." Google Company History, Google, accessed September 18, 2013, <http://www.google.com/about/company/history>.
- ^v "What Is the Core Reporting API – Overview" Google Developers, Google, accessed September 18, 2013, <https://developers.google.com/analytics/devguides/reporting/core/v3>.
- ^{vi} "Dimensions and Metrics," Google Developers, Google, accessed September 18, 2013, <https://developers.google.com/analytics/devguides/reporting/core/dimsmets>.
- ^{vii} "How to Set Up the Web Tracking Code," Google Analytics, Google, accessed September 18, 2013, https://support.google.com/analytics/answer/1008080?hl=en&utm_medium=et&utm_campaign=en_us&utm_source=SetupChecklist.
- ^{viii} "New Acquisitions Reporting & Channels Shows A More Complete Picture Of Your Users & Marketing" Google Analytics, Google, accessed October 9, 2013, <http://analytics.blogspot.com/2013/10/new-acquisitions-reporting-channels.html>

^{ix} “What’s Happening on Your Site Right Now?” Google Analytics, Google, accessed September 18, 2013, <http://analytics.blogspot.com/2011/09/whats-happening-on-your-site-right-now.html>.

^x Tamara Sumner et al., “Developing a Web Analytics Strategy for the National Science Digital Library,” *D-Lib Magazine* 10, no. 10 (2004): 1082-9873, <http://www.dlib.org/dlib/october04/coleman/10coleman.html>.

^{xi} Michael Khoo et al., “Using Web Metrics to Analyze Digital Libraries,” *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries*. (2008): 375-384.

^{xii} Wei Fang, “Using Google Analytics for Improving Library Website Content and Design: A Case Study,” *Library Philosophy and Practice* 9, no. 2 (June 2007): 1-17. <http://works.bepress.com/wfang/1>

^{xiii} Elizabeth L. Black, “Web Analytics: A Picture of the Academic Library Web Site User,” *Journal of Web Librarianship* 3, no. 1 (2009): 3-14.

^{xiv} Anindita Paul, “The Use of Web Analytics on an Academic Library Website,” Doctoral dissertation, University of Missouri-Columbia, 2009.

^{xv} Loftus, “Demonstrating Success: Web Analytics and Continuous Improvement,” 45-55.

^{xvi} Kirk Hess, “Discovering Digital Library User Behavior with Google Analytics,” *Code4Lib Journal* 17 (June 2012). <http://journal.code4lib.org/articles/6942>

^{xvii} Paul Betty, “Assessing Homegrown Library Collections: Using Google Analytics to Track Use of Screencasts and Flash-Based Learning Objects,” *Journal of Electronic Resources Librarianship* 21, no. 1 (2009): 75-92.

^{xviii} Julie Arendt and Cassie Wagner, “Beyond Description: Converting Web Site Usage Statistics into Concrete Site Improvement Ideas,” *Journal of Web Librarianship* 4, no. 1 (2010): 37-54.

^{xix} Steven J. Turner, "Website Statistics 2.0: Using Google Analytics to Measure Library Website Effectiveness," *Technical Services Quarterly* 27, no. 3 (2010): 261-278.

^{xx} Christopher J. Prom, "Using Web Analytics to Improve Online Access to Archival Resources," *American Archivist* 74, no. 1 (2011): 158-184.

^{xxi} Mark R. O'English, "Applying Web Analytics to Online Finding Aids: Page Views, Pathways, and Learning about Users," *Journal of Western Archives Volume* 2, no. 1 (October 2011): 1-12.

^{xxii} Paul Ament-Gjevick, "Using Web Analytics—Archival Websites," (Thesis, Clayton State University, 2012).

^{xxiii} "Bounce Rate," Google Analytics, Google, accessed September 19, 2013, <https://support.google.com/analytics/answer/1009409?hl=en>

^{xxiv} "Back to Basics: Direct, Referral or Organic: Definitions Straight from the Source," Google Analytics, Google, accessed September 19, 2013, <http://analytics.blogspot.com/2009/08/back-to-basics-direct-referral-or.html>.

^{xxv} "Referral Traffic," Google Analytics, Google, accessed September 19, 2013, <https://support.google.com/analytics/answer/1247839?hl=en6>

^{xxvi} "Ralph Satterlee Indianapolis 500 Photographs," Digital Media Repository, Ball State University Libraries, accessed September 18, 2013, <http://libx.bsu.edu/cdm/landingpage/collection/SatterleePh>.

^{xxvii} Jennifer Schaffner, *The Metadata Is the Interface: Better Description for Better Discovery of Archives and Special Collections, Synthesized from User Studies* (OCLC Research, 2009). <http://www.oclc.org/programs/publications/reports/2009-06.pdf>.

^{xxviii} O'English, "Applying Web Analytics to Online Finding Aids: Page Views, Pathways, and Learning about Users," 5.

^{xxix} Prom, "Using Web Analytics to Improve Online Access to Archival Resources," 170.

^{xxx} "Page Tracking—Dimensions & Metrics Reference," Google Developers, Google, accessed September 19, 2013, <https://developers.google.com/analytics/devguides/reporting/core/dimsmets/pagetracking#ga:landingPagePath>.

^{xxxi} Jane Gertz, "Preservation and Selection for Digitization," (Northeast Document Conservation Center), accessed September 19, 2013, <http://www.nedcc.org/free-resources/preservation-leaflets/6.-reformatting/6.6-preservation-and-selection-for-digitization>.

^{xxxii} "The Difference Between Clicks, Visits, Visitors, Entrances, Pageviews, and Unique Pageviews," Google Analytics, Google, accessed September 19, 2013, <https://support.google.com/analytics/answer/1257084?hl=en>.

^{xxxiii} "Content Overview," Google Analytics, Google, accessed September 19, 2013, <https://support.google.com/analytics/answer/2404517?hl=en>.