

Using Web Analytics to Improve Online Access to Archival Resources

Christopher J. Prom

Abstract

Since the late 1960s, archivists and technologists have developed and implemented innovative standards, practices, and technologies intended to facilitate the description, availability, and use of archival materials. Although people can discover, use, and interpret records and manuscripts in new ways, archivists lack a systematic understanding of how people interact with the descriptive information and digital objects that they create and post online. This article introduces Web analytics as a method that archivists can use to measure user actions, to understand some aspects of user behavior, and to initiate a program that will improve online services. By interpreting Web analytics data in light of repository goals and other information concerning use (such as usability studies), repositories can dramatically improve access, increase use, and heighten user satisfaction.

Since the late 1960s, professional working groups, such as the National Information Systems Task Force (NISTF) and its successors, the SAA Standards Committee and Technical Subcommittee on Descriptive Standards, the ICA Committee on Descriptive Standards, the Encoded Archival Description Working Group, and the Encoded Archival Context Working Group, have provided an intellectual and technical framework through which archival descriptive information can be recorded, stored, and exchanged.¹ In

© Christopher J. Prom.

¹ The history of archival descriptive standards is reviewed in Daniel V. Pitti, "Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids," *American Archivist* 60 (Summer 1997): 268–83; Steven L. Hensen, "'NISTF II' and EAD: The Evolution of Archival Description," *American Archivist* 60 (Summer 1997): 284–85; and several chapters in Daniel V. Pitti and Wendy M. Duff, eds., *Encoded Archival Description and the Internet* (Binghamton, N.Y.: Haworth Press, 2001). The development of encoding and data exchange standards continues in the Encoded Archival Context Working Group, which has developed a standard for the exchange of authoritative contextual information regarding the persons, corporate bodies, and families that create archival materials. See EAC-CPF, <http://eac.staatsbibliothek-berlin.de/>, accessed 17 December 2009.

addition, standards developed outside the archival community, such as the Metadata Encoding and Transmission Standard (METS) and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), offer methods to record and share information about digitized surrogates of analog and born-digital records. Guidelines such as *Archives, Personal Papers and Manuscripts*; *Describing Archives: A Content Standard*; and RLG's *Best Practice Guide for EAD* help ensure cross-repository consistency in the format of archival descriptive records.² Numerous software applications provide tools to record and share descriptive information.³ In addition, a wide range of software products, such as CONTENTdm, DSpace, and Fedora, are used to provide access to digital objects.⁴ Archivists and librarians routinely announce technically sophisticated projects using communication tools such as the EAD listserv.⁵ In addition, many archives are implementing so-called Web 2.0 technologies, which promote use through new venues (such as blogs, Twitter, and other social media) or which allow users to personalize or annotate online archival materials. Emerging programming techniques that use so-called semantic Web technologies (i.e., publishing in formats intended for data description and exchange) open up seemingly limitless potential for new

² Steven Hensen, comp., *Archives, Personal Papers, and Manuscripts*, 2nd ed. (Chicago: Society of American Archivists, 1998); *Describing Archives: A Content Standard* (Chicago: Society of American Archivists, 2004). The RLG best practice guide is available at <http://www.oclc.org/programs/ourwork/past/ead/bpg.pdf>, accessed 10 April 2009.

³ Information about some of these tools is available at the following project websites: MicroMARC for Integrated Format, <https://www.msu.edu/user/msumarc/>; "EAD Help Pages—EAD Cookbook 2002," <http://www.archivists.org/saagroups/ead/ead2002cookbookhelp.html>; Archivists' Toolkit, <http://www.archiviststoolkit.org/>; Archon, <http://www.archon.org>; ICA-Atom Open Source Archival Description Software, <http://ica-atom.org/>; and ArchivesSpace, <http://www.archivespace.org>. The Society of American Archivists EAD Roundtable maintains a partial list of other descriptive software tools and projects at "EAD Help Pages," <http://www.archivists.org/saagroups/ead/>. All URLs accessed 31 January 2010. A helpful resource for choosing a descriptive tool is Lisa Spiro, *Archival Management Software: A Report for the Council on Library and Information Resources* (Washington, D.C.: Council on Library and Information Resources, January 2009), <http://www.clir.org/pubs/reports/spiro2009.html>, accessed 17 December 2009.

⁴ See OCLC, <http://www.contentdm.org/>; DSPACE, <http://dspace.org>; FedoraCommons, <http://www.fedora-commons.org/>. KeepSolutions, "RODA" (<http://redmine.keep.pt/projects/show/roda-public>) and University of Virginia Libraries, "Hydra Project" (<http://www2.lib.virginia.edu/innovation/hydra/>) are two Fedora-based preservation and access systems that are being built for the management, preservation, and access of born-digital archival materials. All accessed 11 May 2010.

⁵ For example, see Barbara Aiken's 2 December 2009 announcement regarding the Smithsonian Institution's Collections Search system, ListServ.LOC.gov, <http://listserv.loc.gov/cgi-bin/wa?A2=ind0912&L=ead&T=0&P=56>, or the AIMS project, which seeks to provide a cooperatively developed infrastructure for stewardship of born-digital materials, "AIMS Born-Digital Collections: An Inter-Institutional Model for Stewardship," <http://www2.lib.virginia.edu/aims/>, accessed 2 October 2010.

services.⁶ Open-source libraries and microblogging sites allow for the rapid development of interactive Web resources, presenting archival information in unexpected and innovative ways.⁷

As a result, people are able to discover, use, interpret, and interact with records and manuscripts in ways inconceivable only twenty years ago. Online archival databases, image repositories, and other electronic sources open our archives to new audiences and provide traditional users a way to access, use, and repurpose materials without an archivist's mediation. We can only expect that the future promises further innovation and access enhancements.

All this activity is impressive, but archivists have not developed a systematic understanding of how users interact with the descriptive information posted online. Perhaps this problem might be best expressed as a series of questions: How might we as archivists make archival information optimally accessible to new users? How do we best explain the rich content, structure, and context that constitute our holdings? Are we providing online users what they really want? Do our efforts enhance the experience of on-site and hybrid (online/on-site) users? How can we best optimize our Web pages to encourage user interest and search engine exposure? These questions are important because, as one usability expert explains, "Simply put, if your users have a bad experience, they won't come back."⁸

Many users first experience archival materials through a Web browser. Archival websites function as virtual spaces where users can discover and, increasingly, interact with our collections. Since much of this activity takes place without our knowledge, it is difficult to gather accurate information regarding their online experiences. When we have better information about how people interact with our virtual resources, we can design more effective websites and, we hope, increase our users' satisfaction.

⁶ The World Wide Web Consortium (<http://www.w3.org>) develops and maintains the standards, such as XHTML, SOAP (Simple Object Access Protocol), RDF (Resource Discovery Framework), and OWL2 (Web Ontology Language), that facilitate the development of new access methods.

⁷ In addition, many open-source projects have released rapid development frameworks and libraries that enable computer programmers to develop interactive, layered, and user-friendly access systems. Some of the most popular tools include the Symfony framework, <http://www.symfony-project.org/> and the jQuery library, <http://jquery.com/>. Projects like these often implement the REST architectural style and use AJAX programming techniques, described in more detail at Roger L. Costello, "Building Web Services the REST Way," xFront, <http://www.xfront.com/REST-Web-Services.html> and W3Schools .com, "AJAX Introduction," http://www.w3schools.com/Ajax/ajax_intro.asp. Additional examples of such tools can be readily identified using the *Wikipedia* pages describing REST and AJAX. All accessed 1 October 2010.

⁸ Jesse James Garrett, *The Elements of User Experience: User-Centered Design for the Web* (New York: AIGA/New Riders, 2003), 14.

Archival user behavior, per se, has not been a neglected topic of research—although this accusation has been made even in the fairly recent past.⁹ For many years, archivists gathered reference statistics, consulted with the users of archival records and manuscript collections, and conducted both formal and informal user studies. The archival literature includes a wide range of research and informed opinion concerning user research needs, preferences, styles, and behaviors. This literature utilizes cogent methodologies such as surveying or interviewing current users, analyzing reference correspondence or email, and conducting user surveys or interviews.¹⁰

However, few of these studies discuss the specific ways that users interact with online archival resources and services, and even fewer employ a rigorous methodology based upon stated evaluation criteria.¹¹ Most of our knowledge about user online behavior is drawn from survey evidence or has been gathered in the confines of the physical archives under a controlled setting.¹² While these methods are useful, other options are needed.

Web Analytics Overview

Web analytics is a new method that archivists can use to measure user actions, to understand some aspects of user behavior, and to initiate a program to continuously improve online services. According to the Web Analytics

⁹ For example, see Wendy M. Duff, Jean Dryden, Carrie Limkilde, Joan Cherry, and Ellie Bogomazova, "Archivists' Views of User-based Evaluation: Benefits, Barriers, and Requirements," *American Archivist* 71 (Spring/Summer 2008): 145, which cites several examples of calls for increased user study and evaluation. In addition, Paul Conway's recommended for a suite of user studies (most of which have not been completed to my knowledge) in "Facts and Frameworks: An Approach to Studying the Users of Archives," *American Archivist* 49 (Fall 1986): 393–408. The Archival Metrics Project provides a standard set of user-based evaluation tools that offer the possibility of comparative research and understanding of user needs across repositories, <http://www.archivalmetrics.org/>, accessed 17 December 2009.

¹⁰ Classic studies of exceptional value include William J. Maher, "The Use of User Studies," *Midwestern Archivist* 11, no. 1 (1986): 15–26. More recent works of note include Wendy Duff and Catherine A. Johnson, "A Virtual Expression of Need: An Analysis of E-Mail Reference Questions," *American Archivist* 64 (Spring-Summer 2001): 43–60; Susan Hamburger, "How Researchers Search for Manuscript and Archival Collections," *Journal of Archival Information* 2, nos. 1–2 (June 2004): 79–102; and Elizabeth Yakel, "Managing Expectations, Expertise, and Effort While Extending Services to Researchers in Academic Archives," in *College and University Archives: Readings in Theory and Practice*, ed. Christopher J. Prom and Ellen D. Swain (Chicago: Society of American Archivists, 2008), 261–86.

¹¹ Duff et al., "Archivists' Views of User-based Evaluation," 145–47. See Morgan G. Daniels and Elizabeth Yakel, "Seek and You May Find: Successful Search in Online Finding Aid Systems," *American Archivist* 73 (Fall/Winter 2010): 535–68.

¹² Studies that comment specifically on understanding user needs in the online environment or on designing effective online services include Burt Altman and John Nemmers, "The Usability of On-line Archival Resources: The Polaris Project Finding Aid," *American Archivist* 49 (Fall 1986); Elizabeth Yakel, "Encoded Archival Description: Are Finding Aids Boundary Spanner or Barriers for Users?," *Journal of Archival Organization* 2, nos. 1–2 (2004): 66–37; Christopher J. Prom, "User Interactions with Finding Aids in a Controlled Setting," *American Archivist* 67 (Fall/Winter 2004): 234–68; Katherine A. Salzmann, "'Contact Us': Archivists and Remote Users in the Digital Age," *Reference Librarian* 85 (2004): 43–50.

Association, Web analytics is “the measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage.”¹³ In the past it was also defined as “the study of the behavior of website visitors.”¹⁴ *Wikipedia* adds, “[I]n a commercial context, Web analytics especially refers to the use of data collected from a website to determine which aspects of the website work towards the business objectives; for example, which landing pages encourage people to make a purchase.”

There is nothing specifically archival about Web analytics. It is a business tool developed to serve a commercial purpose. Online businesses maximize profit when they design websites that make it easy for people to buy goods or services. Web analytics software helps online businesses improve their websites and advertising campaigns, so that people buy more goods or services.

Many of the usability techniques used by businesses can and should be adopted by libraries and archives. Leslie Porter focuses particular attention on process-oriented testing, side-by-side comparison, and task simulation (such as buying a small item using real money) as business methods that libraries can use to approximate real-life experiences.¹⁵ Web analytics is another such method, but surprisingly little attention has been paid to its capabilities for libraries or archives. A comprehensive literature search revealed only two articles, which discussed how the specific features included in Web analytics software can be adapted to help meet the business objectives of a library.¹⁶ No published articles discuss its potential use on the websites of archival repositories, although the Association of Research Libraries (ARL) offered a workshop about it.¹⁷

We can use analytics tools to improve our websites, to make effective decisions regarding online services, and to improve our users' experiences. In other words, the tools and concepts that help businesses understand user needs and behaviors to maximize profit can be used to help archives understand user needs and behaviors to maximize archival use.

¹³ Web Analytics Association, “About Us,” <http://www.webanalyticsassociation.org/?page=aboutus>, accessed 18 October 2010.

¹⁴ “Web analytics,” *Wikipedia*, s.v. http://en.wikipedia.org/wiki/Web_analytics, accessed 19 November 2007.

¹⁵ Leslie Porter, “Library Applications of Business Usability Testing Strategies,” *Library Hi Tech* 25, no. 1 (2007): 126–35.

¹⁶ Marshall Breeding argues that unlike traditional methods of studying users, Web analytics offer the ability to analyze data that is based on actual use, based on the ideas that “on the Web, each keystroke and mouse click—or its absence, registers as a vote” and that cumulatively, this information “describes the site’s overall usability and usefulness.” Marshall Breeding, “An Analytical Approach to Assessing the Effectiveness of Web-based Resources,” *Computers in Libraries* 28 (January 2008): 20–22. Feng Wei provides an overview of Google Analytics and illustrates how the Rutgers University Law Library used it to understand users’ computer capabilities (e.g., screen resolution) to track resource usage (e.g., most popular pages) and to facilitate a decision-making process to improve website design. Feng Wei, “Using Google Analytics for Improving Library Website Content and Design: A Case Study,” *Library Philosophy and Practice* (July 2007), <http://digitalcommons.unl.edu/libphilprac/121/>, accessed 13 November 2009.

¹⁷ Association of Research Libraries, “Statistics and Assessment,” <http://www.arl.org/stats/statsevents/google-analytics-workshop.shtml>, accessed 1 June 2010.

Improving website usability needs detailed, accurate, and useful data, but reporting mechanisms such as server logs often provide rudimentary, misleading, or irrelevant data.¹⁸ Such data cannot serve as the basis for a decision or action. For example, internal server logs at the University of Illinois Library report only how many “hits” were recorded, not how long a user stayed on a page or how many other pages he or she visited. Other methods are necessary to represent actual user behavior.

In contrast to internal server logs, Web analytics applications provide accurate and deep information about the actual use of online resources. For example, they report which websites referred users to the site, how long users stayed on particular pages, and how many pages they viewed during a visit. Their reports exclude traffic generated by nonhuman agents (such as Web crawlers), so only actual use is included in the reports. Analytics software can also help an archivist understand how a particular resource type is being used. For example, the software can measure the frequency with which a page is viewed after people view the page’s parent record.

Web analytics software has another advantage: It provides a means to indirectly observe some of the ways in which users interact with online resources while they are seeking information relevant to an actual research need. Most other methods of studying user behavior require intervention, contact, or observation, each of which inevitably affects user actions. For example, many user studies ask people to complete a canned search or to “act normal,” even though they are being observed or even videotaped. Of course, Web analytics cannot substitute for consulting users or conducting formal usability studies. Used in conjunction with other, more traditional methods of studying user behavior, however, it can force us to ask new questions about users and their information-seeking behaviors.

Web analytics includes two discrete elements: using a software tool to collect, measure, and report user data; and interpreting the reported information to make decisions regarding services to improve a website’s features. The reports that the software generates facilitate effective analysis and decision making since the types of information reported are more discrete and detailed than others to which archivists might have access, such as server logs. Web analytics come in two basic flavors: log-based and script-based services. Log-based services, such as WebTrends, parse data that is natively collected by server logs. Script-based

¹⁸ For example, internal logs analysis conducted at the beginning of this study found that the University of Illinois Archives received 803,838 hits in July 2007, but only a small percentage represented human use. Server logs typically track hits. However, a hit is simply a request from another computer to a Web server. As a data point, the concept of the hit is inherently problematic for several reasons. Hits originate from both humans and from nonhuman agents, such as a search engine’s Web crawler. In addition, a single page request is often recorded as several hits since each embedded image, is often recorded as a separate hit. Thus a page with twenty-five embedded images records twenty-six hits each time it is accessed, and the number of hits does not accurately represent human interaction with the website.

services, such as Google Analytics, report information to a third-party server each time a page is requested. Google Analytics is readily accessible to archives since it is free and can be installed with little or no technical know-how, but it is important to note that script-based services can underreport the actual use of the site. For example, they will not provide information for Web users who have disabled third-party cookies using their browsers' privacy settings.¹⁹

In spite of this drawback, script-based analytics services are very useful for assessing user behavior. Properly implemented, they provide data that complement other methods of studying user behavior and needs, such as usability tests or human-computer interaction (HCI) studies.²⁰ The proper use of Web analytics may help a repository develop an effective program of in-person usability tests. It can also help institutions assess whether particular changes made to a website have had a positive effect on website usage and user satisfaction. For example, if a repository reformats its digital object delivery page based on user feedback, analytics software will reveal whether digital objects were downloaded at a higher relative frequency after the change—an indirect measure of page impact and, presumably, user interest.

In addition, Web analytics can be implemented as a stand-alone process (as described in the remainder of this article) to understand user behavior and improve archival websites. Such a strategy can be effective if an institution does not have the resources or inclination to develop a full usability program. The method in which we implemented a Web analytics program at the University of Illinois illustrates this approach and has helped us develop a baseline for other usability studies.

¹⁹ The command to submit information to the third-party server is executed by Javascript code in the header or footer of each page. It would appear likely that the percentage of such users is low, but no recent statistics exist. In 2000, it was reported that 10 percent of all Internet users blocked cookies, but doing so in 2010 would likely make the Internet all but unusable for most people, since the practice of setting cookies is so prevalent. Susannah Fox, *Trust and Privacy Online*, Pew Internet and American Life Project (2000), <http://www.pewinternet.org/Reports/2000/Trust-and-Privacy-Online.aspx>, accessed 31 January 2011. In addition, users may have strict antivirus settings that do not allow the embedded script to run, or they may be browsing in private mode, so that they do not leave a trace of their activity. Apparently, Google now allows people to opt out of analytics tracking altogether. Taken as a whole, these factors lead to underreporting of actual website use. For more information, see *Google Analytics*, "More Choices for Users: Browser-Based Opt-Out for Google Analytics on the Way," <http://analytics.blogspot.com/2010/03/more-choice-for-users-browser-based-opt.html>, accessed 30 September 2010.

²⁰ Numerous books and resources discuss how HCI studies should be conducted or provide effective case studies. Useful planning guidance and case studies include Garrett, *The Elements of User Experience*; Brenda Battleson, Auston Booth, and Jane Weintrop, "Usability Testing of an Academic Library Web Site: A Case Study," *Journal of Academic Librarianship* 27 (May 2001): 188–98; and Chrysanthia Malama, Monica Landoni, and Ruth Wilson, "What Readers Want: A Study of E-Fiction Usability," *D-Lib Magazine* (May 2005) <http://www.dlib.org/dlib/may05/wilson/05wilson.html>, accessed 30 September 2010. The website of the Society for Technical Communication's Usability and User Experience Community also includes a wide range of resources, including a Usability Toolkit, that are helpful when planning and implementing usability tests and user interviews, <http://www.stcsig.org/usability/>, accessed 30 September 2010.

Project Preparation and Methodology

In early 2007, the University of Illinois Archives implemented a pilot project to test the utility of Google Analytics for measuring and analyzing use of our website.²¹ We hoped that the project would provide preliminary information about online use and demonstrate whether the concept of Web analytics might have broader applicability in the archival community, both in the United States and internationally. As the project progressed, it became apparent that it would allow us to identify usability impediments and to iteratively improve our website, better meeting our revised understanding of user needs. In this sense, we used Web analytics as a method to implement principles of user-centered design by making small progressive changes to an existing website over a period of several years.²²

Before beginning the project, we analyzed existing data about the use of our archives, developed a set of questions to examine during the pilot project, and developed a privacy policy. These steps helped ensure that the information we gathered and analyzed was accurate and useful.

The University of Illinois Archives has gathered reference statistics since its establishment in 1963, and William Maher has described use trends at Illinois through 1985.²³ While this article is not the appropriate forum to update that analysis, it is clear that use has been both increasing and migrating from on-site to remote services. In our fiscal year 2005–2006, 782 (28%) of our 2,782 user contacts were initiated via email. An additional 367 (13%) were initiated via telephone. Relatively few of these users subsequently visited the archives. We provided them a variety of remote services ranging from photocopying and scanning to complimentary and fee-based research. Using the reference statistics system that produced these figures, it is impossible to gather any information regarding those who did not contact us but used our website in some fashion. Therefore, it seemed reasonable to use the analytics software to understand how all of our users were interacting with the site so that we could make changes, increase their satisfaction, and (we hoped) use.

²¹ "Google Analytics," <http://www.google.com/analytics>, accessed 30 September 2010.

²² According to an authoritative definition, "User Centered-Design (UCD) is a philosophy and a process. It is a philosophy that places the person (as opposed to the 'thing') at the center; it is a process that focuses on cognitive factors (such as perception, memory, learning, problem-solving, etc.) as they come into play during peoples' interactions with things." Raïssa Katz-Haas, "Usability Techniques: User-Centered Design and Web Development," Society for Technical Communication, http://www.stcsig.org/usability/topics/articles/ucd%20web_devel.html, accessed 30 September 2010. Other useful resources that discuss user-centered design principles include Nancy A. Van House et al., "User-Centered Iterative Design for Digital Libraries: The Cypress Experience," *D-Lib Magazine* (February 1996), <http://www.dlib.org/dlib/february96/02vanhouse.html>, accessed 1 October 2010; and Steve Krug, *Don't Make Me Think: A Common Sense Approach to Web Usability*, 2nd ed. (Berkeley, Calif.: New Riders Publishing, 2006).

²³ Maher, "The Use of User Studies."

Assessment of website use must be conducted against objectives that measure, in a statistically accurate way, whether the purposes or goals of the website are being achieved. The website of the University of Illinois Archives serves several complementary purposes. Specifically, it provides basic information about the archives (such as contact information and a general description of our holdings), promotes our programs and services, encourages users to find relevant descriptive information, and facilitates use of online and offline materials. It would be tempting to analyze information regarding all of these goals. However, anyone conducting a user study should be reluctant to analyze a large number of data points, since many will likely be irrelevant to the task of understanding user behaviors and usability problems. It makes more sense to analyze closely defined data that speak to a well-defined hypothesis. The Web measurement literature notes that a successful analytics program will collect a limited amount of data and will seek to answer a few discrete and measurable questions based on an analysis of Web traffic over a defined period of time.²⁴ Web analytics software can be configured to provide specific types of information if you have precise questions in mind.

Based on our use trends, we felt it was most important to find out whether our website's design encouraged or discouraged archival use and contacts with archival staff. Not only are such contacts reported to our administration as evidence of our use and effectiveness, but they also represent one way of measuring whether our online services have an impact on our users. We developed a hypothesis and four specific questions intended to measure whether our hypothesis was true. We hoped that by analyzing the data around these specific questions, we would not only prove or disprove the hypothesis, but identify specific actions that we could take to improve the website and make decisions regarding future services.

The hypothesis was that typical users would enter our site through our homepage, complete a search, look at the holdings records, view the full finding aid, and—if they felt the records described in the finding aid might meet their information need—contact us for more information. I tested this hypothesis using four questions:

- 1) Which parts of our website are most heavily used?
- 2) How do people reach our site?
- 3) What are the most popular searches on our site?
- 4) How do users navigate through our website (and in particular, which pages lead them to contact us)?²⁵

²⁴ Eric T. Peterson, *Web Site Measurement Hacks: Tip Tools to Help Optimize Your Online Business*, (Sebastopol, CA: O'Reilly Press, 2005), 54–55, 72.

²⁵ In later stages of an analytics program, it would be better to have a more precise question in mind such as "Which pages lead users to contact us?" However, attempting to measure such a precise goal early in the study was deemed useless, since it would be impossible to identify large-scale usability problems until developing a broad understanding of how users move through the website.

Installing and Using the Web Analytics Software

Those interested in Web analytics can select from many vendors or services.²⁶ For archivists, Google Analytics is probably the most feasible option not only because it is free and easy to install, but because it reports very useful data.

Before attempting to install the software and configure it, we developed a privacy policy regarding the use of the analytics program, working with the campus IT officer. Not only did state law require such a policy, but it was also good ethical and professional practice to inform users that we and Google were collecting and analyzing anonymous information concerning their visits. We also provided instructions so that individual users could disable tracking.²⁷ Developing the policy forced us to think clearly about the information we were gathering and how it would be used.²⁸

After the privacy policy had been approved and posted, we registered an account at www.google.com/analytics and loaded tracking code (a reference to a javascript library) onto every page on our website. For the most part, this was a simple process because many pages on our site include a common header and footer.²⁹

After installing the software in June 2007, I configured it to ensure that it collected data that would be accurate and relevant to our hypothesis and research questions. Tools built into the Google Analytics software provided data that can answer the first three questions. To gather data regarding the last question, I defined certain user behaviors that seemed desirable, such as viewing a descriptive record or emailing the archives, then I configured the analytics software to report how and when users exhibited the desired behaviors. In practice, this required that “goals” be defined in the analytics software, so that progress toward the goals could be measured.

The idea of measuring progress toward a goal is predicated upon the analytics concept of the “goal conversion funnel.” In a commercial setting, a business

²⁶ *Wikipedia* includes a list of the most popular vendors at http://en.wikipedia.org/wiki/List_of_web_analytics_software, accessed 11 May 2010.

²⁷ While Google appears to collect and store IP addresses as part of the tracking process (as do most internal server logs), IP addresses are not provided to end users of the analytics reports. It is subject to the “Google Analytics Terms of Service,” <http://www.google.com/analytics/tos.html> and to the Google privacy policy, Google Privacy Center, “Transparency and Choice,” <http://www.google.com/intl/en/privacy/>, both accessed 31 January 2011. There is some evidence that Google anonymizes IP addresses after nine months. Nate Anderson, “Why Google Keeps Your Data Forever, Tracks You with Ads,” *Ars Technica, Law and Disorder* 9 (9 March 2010) <http://arstechnica.com/tech-policy/news/2010/03/google-keeps-your-data-to-learn-from-good-guys-fight-off-bad-guys.ars>, accessed 1 October 2010.

²⁸ A copy of our privacy policy is available at <http://www.library.illinois.edu/archives/about/privacy.php>, accessed 11 May 2009.

²⁹ For pages that did not include the header, we used a directory-level “find and replace” to insert the code, using Note Tab Pro.

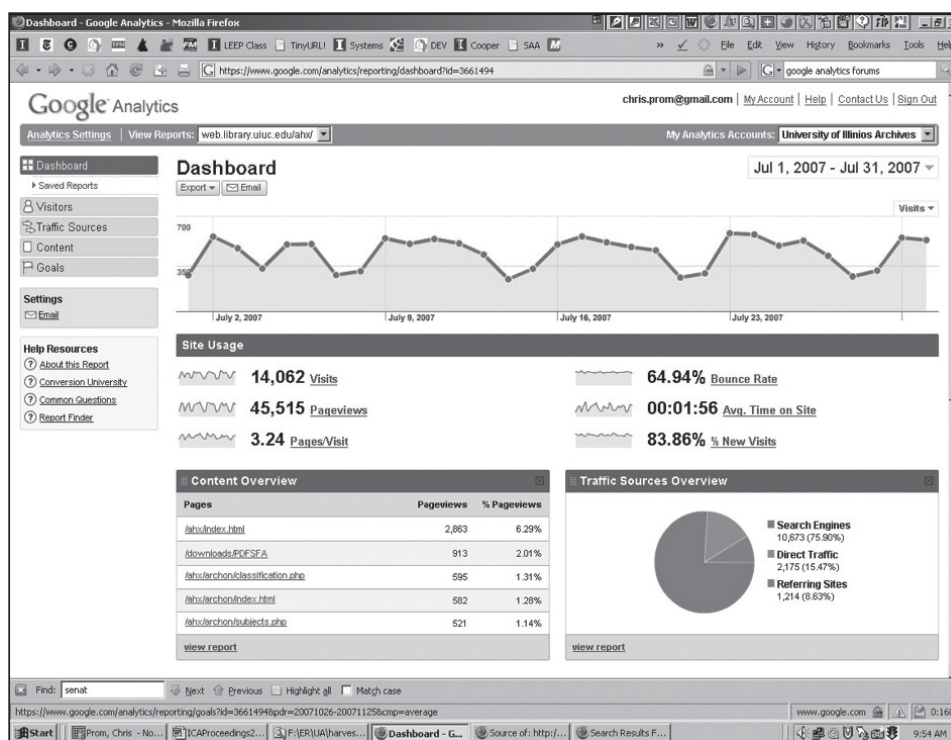


FIGURE 1. Google Analytics dashboard, August 2007.

would know that a goal has been achieved when the browser displays a thank-you page after a user purchases something. As users (in an aggregate sense) progress toward the goal, one can imagine them being winnowed. At each step in a process (such as searching for an item, adding it to a basket, entering payment information, or completing the sale), a certain number of users will abandon the process or leave the site.³⁰ Obviously, the number of users who reach the goal (in both absolute and relative terms) will determine the success or failure of a business. If a large number of users drop off at a particular step, the page in question likely has usability problems or is not meeting user needs.

The University of Illinois Archives defined goal conversion funnels that we believed would help answer my fourth research question: “How do users navigate through the site?” One desirable user behavior seemed to be navigating to

³⁰ Google Analytics concisely defines a funnel as “a series of pages through which a visitor must pass before reaching the goal conversion. The name comes from a graph of visitors who reach each page—the first page counts the most visitors, and each successive page shows less [*sic*] visitors as they drop off before reaching the final goal,” Google Analytics, “What Is a Funnel?,” <http://www.google.com/support/analytics/bin/answer.py?hl=en&answer=55594>, accessed 30 September 2010. The concept of the funnel is useful in helping to analyze how efficiently the website directs visitors toward a website goal. Pages that are not user friendly will see higher drop-offs.

the page in our database application that shows a series-level descriptive record, then using an email link to contact the archives or get other contact information. In addition, I specified goals to measure whether other desirable behaviors were being completed, such downloading a full finding aid in PDF format.

After we verified that data was being accurately received and reported, we viewed reports using the customizable Google Analytics “dashboard,” shown in Figure 1. The dashboard provides basic data, such as the number of visits and visitors, the top “referrers” (i.e., the websites that users visited immediately prior to viewing our page), and the most-viewed pages. Each report can be expanded (“drilled” in analytics parlance) to provide specific information regarding a single page or group of pages. For example, I mined the reports in the content area to determine which Google searches referred the most users to our site and to see which page types caused users to leave our site in the greatest relative numbers.

Interpreting Analytics Reports

One must be very careful in making changes to a website, much less to an archival program based on one source of data. Nevertheless, a targeted analysis of the results returned by the Google Analytics report tools yielded several interesting findings regarding how users interact with our site. Initially, we analyzed data from website use in July 2007 and discovered information that spoke to each of our four research hypotheses. After interrogating this data, we identified likely user impediments, redesigned pages to remove the impediments, and made informed decisions to allocate resources toward augmenting particular sections of the website.

Question 1: Which Parts of Our Website Are Most Heavily Used?

This question was answered by the drilling and filtering information found in the Content Drilldown section of the Google Analytics dashboard. By navigating the hierarchy, I was able to determine that 63.2% of the “pageviews” on our site during July 2007 were in our Archon holdings database, which provides our finding aid system and access to selected digital content.³¹ This suggested that

³¹ July 2007 provides the most reasonable basis for comparison because parts of the website were moved to a new subdomain in August 2008, inadvertently causing tracking failures for a significant number of pages on our site for the period from August 2007 to February 2008, including all pages generated by the database. Accurate reporting began again 1 March 2008, and subsequent analysis of analytics reports for March, April, and May of 2008 showed broadly similar data to that gathered for July 2008. For example, 67.5% of all page visits in that three-month period were to the Archon holdings database.

initial attempts to improve the usability of this site should focus on that particular resource (see Figure 2).

Analyzing how people found the site confirmed this initial impression and helped identify specific pages needing improvement.

Question 2: How Do People Reach Our Site?

The Google Analytics Traffic Sources Overview provided information regarding this question. As shown in Figure 3, 76% of the approximately 14,000 visits to our site in July 2007 originated from a search engine result page (nearly always Google).³² By contrast, 15.7% of the visits originated when a user entered the URL directly or was placed there when a browser opened the site as the homepage), and 8.6% began when a user clicked a link on another site, such as *Wikipedia* or our parent institution's website.

While the fact that most of our traffic originated with Google was hardly shocking, the results seemed to verify that our descriptive records were being indexed and that search engines were driving many users to our site. As we drilled down into the results, we discovered information that was more unexpected and useful. For example, the approximately 10,000 Google users who

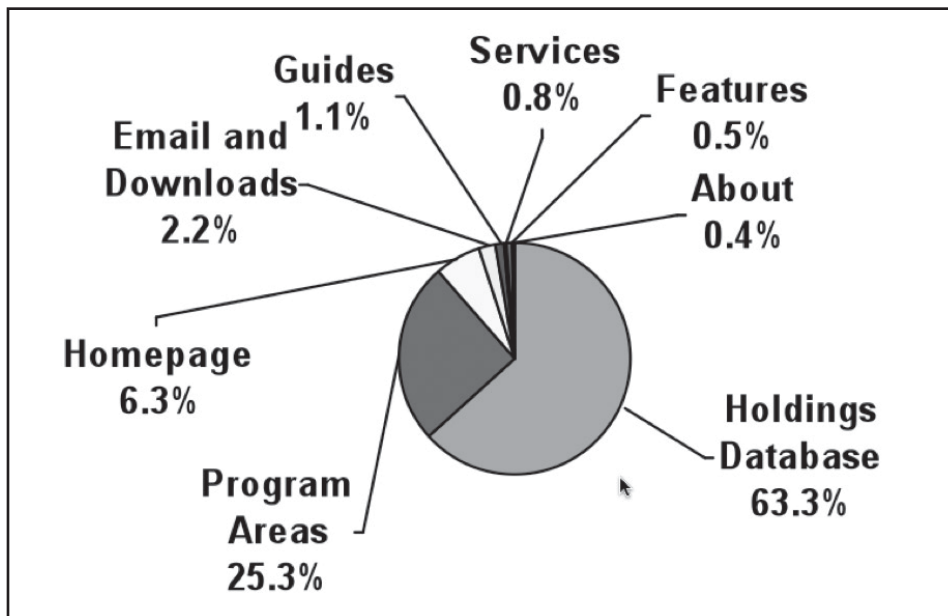


FIGURE 2. Page visits by functional area of website, July 2007.

³² Subsequent analysis showed a drop-off in the number of referrals from search engines; for the period from August 2007 through January 2008, 53.7% originated from them; for July 2009 through June 2010, 57.35% did so.

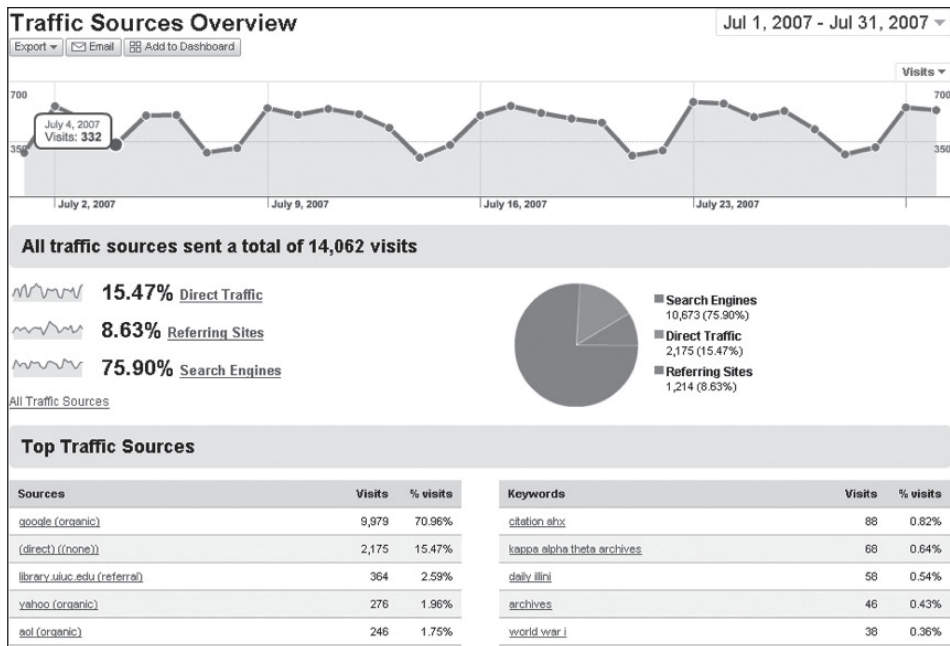


FIGURE 3. Traffic sources overview as of August 2007.

landed on our site in July 2007 stayed on the site for an average of 1 minute 9 seconds before leaving.³³ Only 49 people used the email links to contact us.³⁴ Seventy percent of visitors bounced from the site after viewing only one page.³⁵ It could be argued that few if any of these users seemed to be finding the information they were seeking or contacted us for follow-up information. Nevertheless, 1,977 users did look at additional pages after landing on our site, but only 49 of them tried to contact us by clicking on a link to our email form/address.³⁶ These results indicated that we must determine which pages Google users were being directed to and then make sure that those pages were optimally structured.

³³ This figure does not necessarily represent the actual time users spent on the site, since Google calculates it by comparing a time stamp generated when a user enters the tracked site with a second time stamp generated when they enter another site. Computers that sit idle on our site therefore skew the results upward.

³⁴ Installing the analytics software made us aware that we did not have the means to measure how many contacted us via phone or in person after using the website.

³⁵ This figure is high when compared to nonarchival content sites. One well-regarded analytics blog notes that "Content websites with high search visibility (often for irrelevant terms) can bounce at 40–60%." *Blackbeak's Blog...All Things Analytics*, "Bounce Rate or Single Page Access Industry Averages," <http://blackbeak.conversionchronicles.com/2006/04/12/bounce-rate-or-single-page-access-industry-averages/>, accessed 12 May 2009.

³⁶ The exact number of users who completed the process by actually mailing is not known, since it was not tracked at the time the data was gathered.

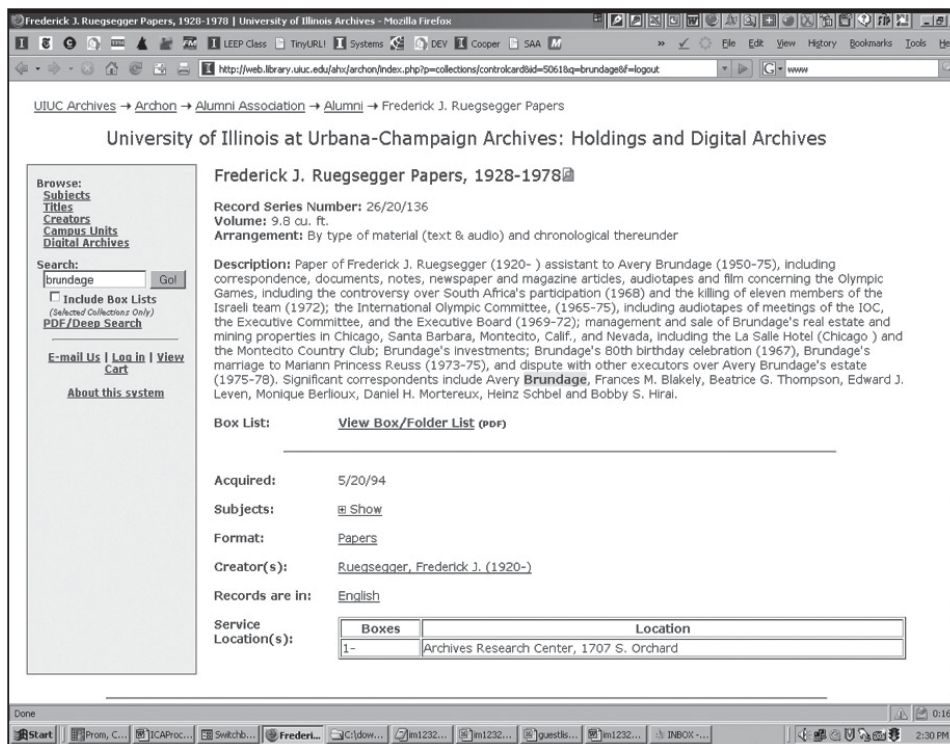


FIGURE 4. Typical landing page, July 2007.

Optimization would encourage users to stay on the site or contact us for additional help.

By accessing the Top Landing Pages report and applying filters to the search result, we determined that 11.9% (1,674) of the website visits began through our homepage, but we surmised that a large percentage of these were staff users.³⁷ In any case, far fewer users see the home page than we hypothesized. By contrast, 47.8% of our users entered our site on one of the over 5,500 series-level description pages generated by our holdings database system. An additional 7.5% of users landed on other pages generated by our database, so over 55% of all visitors to our site in July 2007 first saw something generally similar to the page illustrated in Figure 4. Subsequent analysis showed a similar figure for March through May 2008: 50.8%. Our efforts to improve our site needed to focus first on these pages.

While these landing pages had a few things to recommend them (in particular their simplicity), we surmised that reformatting them slightly would

³⁷ Most views of the homepage are probably accounted for by archives staff computers, which have the page set as home in their browsers, since 1,018 visits came directly to this page via a new browser session.

improve users' experiences and encourage them to further explore our site or contact us for assistance. Further analysis confirmed this impression.

Question 3: What Are the Most Popular Searches on Our Site?

An analysis of analytics data relevant to question 3 helped us better understand some usability problems related to the pages illustrated in Figure 4 and in others in the database's public interface. Some of the issues that we identified are mere usability problems, which could be addressed by reformatting existing data. However, the real value in analyzing the search results was that it demonstrated areas where the content of records needed to be improved or augmented to better aid user understanding and meet informational needs.

Currently, Google Analytics provides two methods to understand how users search for information related to a website: an analysis of search terms entered on search engines embedded in the site (in analytics parlance, "site search"), and an analysis of Google search queries ("keywords") that lead users to the site. Of these, only keyword analysis was available for July 2007, since site search had been recently made available and was not configured for our site until a later date.

To replicate users' experiences and diagnose usability problems, we developed a simple method: We identified common search terms, then replicated each search in Google to mimic the user's interaction with our site. In general, the process of rerunning these searches revealed that many of the visitors who landed on our site ran into an informational dead-end.

For example, 37 users entered our site on the page describing the scrapbook of alumnus Clara Hamilton (see Figure 5). However, 36 bounced immediately from this page and back to the Google search results. Since this page was the first one linked in a Google search for "Clara Hamilton," we could surmise that those users did not find the information they were seeking. Perhaps it was the wrong Clara Hamilton, or perhaps they simply wanted actual information about her, not a description of a scrapbook. Whatever the case, the mere act of rerunning the search demonstrated that the landing page would probably be disorienting to a user, since it does not provide much visual interest or even a biography.

Another pattern emerged from the landing page analysis: The pages typically did not provide archival context information (information about the creator) in a prominent location on the page. Indeed, we had never written biographical notes for many creators, making it difficult for Internet users to decide if the materials would be valuable to their research. These problems led us to

UIUC Archives → Archon → Student Affairs → Student Scrapbooks and Papers → Clara Hamilton Scrapbook

University of Illinois at Urbana-Champaign Archives: Holdings and Digital Archives

Clara Hamilton Scrapbook, 1943-1946

Record Series Number: 41/20/180
Volume: 0.2 cu. ft.

Description: Scrapbook of Clara Hamilton (1923 - ?) (PhD. Zoology, 1943-46) contains photos of individuals and groups of individuals, herself, the University of Illinois campus and buildings, a news clipping, four postcards, and a University of Illinois Weekly Calendar, May 19-26, 1946.

Box List: None available online. [Contact UIUC Archives' staff for assistance.](#)

Acquired: 10/10/2006

Subjects: None currently listed

Creator(s): [Hamilton, Clara \(1923 -\)](#)

Service Location(s):

Boxes	Location
1	Archives Research Center, 1707 S. Orchard

Browse:
Subjects
Titles
Creators
Units
Digital Archives

Search:

☐ Include Box Lists
(Selected Collections Only)

[PDF/Deep Search](#)

[Contact Us](#) | [Log In](#) | [View Cart](#)

[About this system](#)

FIGURE 5. Series-level descriptive record for Clara Hamilton.

suspect that not only did the page need to be optimized, but that it might be wise for us to increase the amount of contextual information, such as biographical notes, provided for our manuscript collections, subject to the availability of resources for retrospective work.

Closer examination of the analytics report also led us to conclude that we needed to integrate digital content more tightly into our database records. A Google search for the term “daily illini” (the title of the campus newspaper) led 58 users to our site in July 2007. Replicating this search in Google confirmed that users were being directed to dead-end descriptive records in our database, since no direct link was made to digital content, even though selected years of the paper had been digitized. As a result, users left the site quickly. This finding motivated us to allocate resources toward a project to tie our digital content delivery mechanism closely to our records-description database.

By analyzing search results from some of the less-popular searches, we identified another major usability problem with the database. For example, one user searched for the phrase “strip mines and Illinois” in Google. The user entered our site, looked at 82 pages over a 28-minute period, and then left for another site or closed the browser.³⁸ Rerunning the search in Google illustrated the problem: The user clicked a vaguely titled link in Google’s results page, reading “Search Results: University of Illinois Archives,” then landed on a page in our database that did not highlight the relevance of the search results to the user (see Figure 6).

The link in Google provided an ambiguous description of the page content. To make matters worse, the landing pages from the Google links represented summarized, collapsed search results. Not only would these problems affect the Google link for this page, but also that for each of the over 4,000 subject terms

³⁸ Our reference statistics show that no users contacted us offline regarding that topic in the month, so it seems safe to assume that the user did not find what he or she was looking for.

in our controlled vocabulary. We resolved to fix the script that generated this page so that it would provide more accurate metadata in the HTML <title> element (which is displayed by Google), then display the results page in a much more cogent fashion.

Question 4: How Do Users Navigate through Our Site?

We attempted to answer this question by tracking user progress toward four goals. Specifically, we tracked how often users completed searches in our holdings database, viewed our series-level descriptions (known informally at the University of Illinois Archives as "control cards"), downloaded a full finding aid (in PDF format), and clicked a link to send us an email message.

Few if any of the people using our website conformed to the linear funnel that I envisaged for the typical user. During July 2007, users viewed our control card descriptive records 8,035 times and downloaded PDF finding aids 697 times. However, only 49 of these views resulted in an email to the archives, suggesting very large fall off in the goal conversion funnel and a need to revise the page to make the "email us" link more prominent.

In general, results from the initial goal tracking were disappointing since we encountered user behavior that did not match our predefined hypothesis. In particular, two of our goals (viewing a complete descriptive record or viewing a search results page) were too broadly defined to yield meaningful data. Users entered these pages from outside our website, and the simple delivery of the page did not really demonstrate a desired behavior. Actions that they might undertake after this point, such as emailing us or downloading digital content, would be better measures of whether we had succeeded in piquing interest or

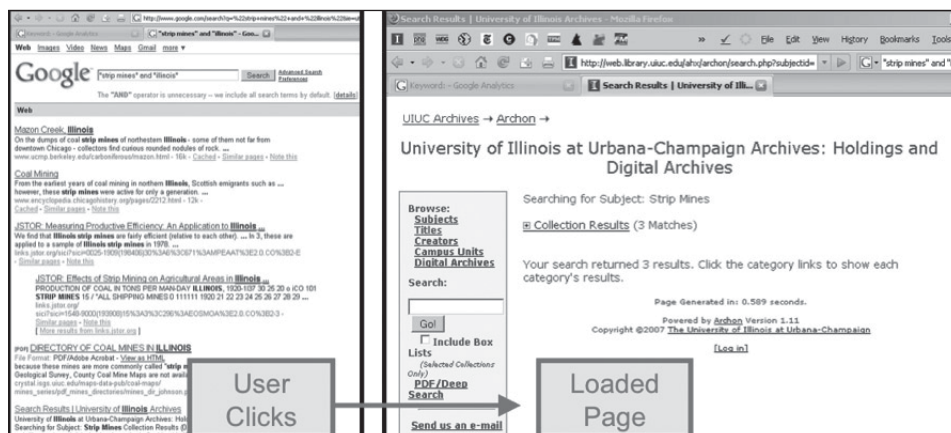


FIGURE 6. Ambiguous link in Google and confusing landing page.

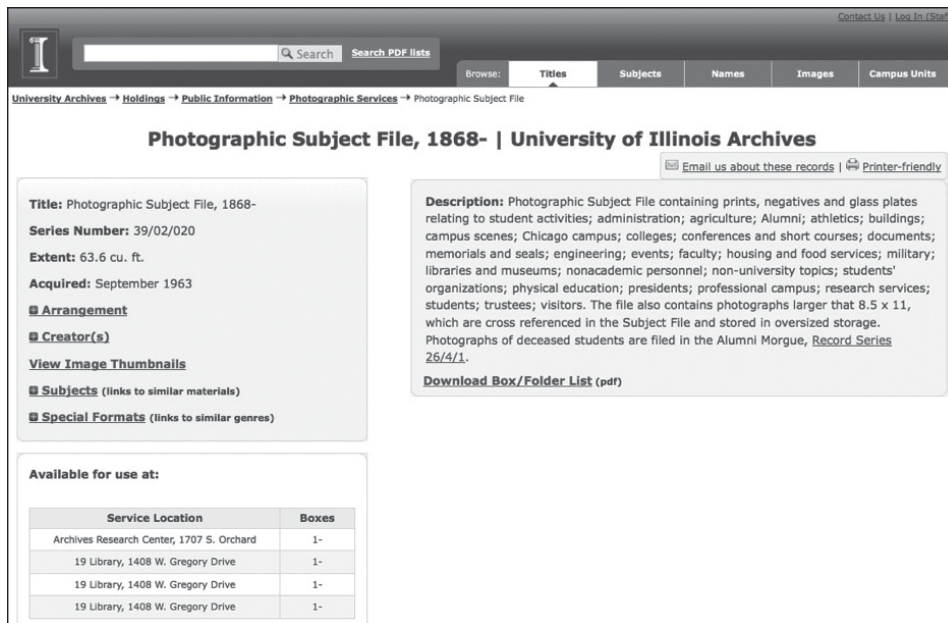


FIGURE 7. Redesigned series-level description view, July 2009.

meeting a need. In this respect, we needed to define a new goal for subsequent tracking: whether a user downloaded digital content.

Implementing Changes Based on Data Analysis

Based on the results of the pilot study, we made two fundamental decisions regarding our website and the Archon database in particular. We also revised our analytics reporting mechanism so that future data would be more cogent and provide information to drive further website improvements.

First, we redesigned the most common landing page (our series-level descriptive records) to emphasize the description of the materials by placing it in a sweet spot based on the design principle known as the “rule of thirds.” We moved the links to our email form, associated digital content, and detailed finding aids closer to the scope and content (and the visual sweet spot), hoping that these links would be utilized more often. We moved all content higher on the page and added information regarding where the materials were physically located. We segregated the search box from the series description, hoping to provide a better visual clue that the search scope would target additional records (not just those described on the landing page). Finally, we moved the browsing options to the top of the page and branded the entire site to make it obvious

that the resources were located at the University of Illinois. The redesigned series-level landing page is shown in Figure 7.

Second, we decided to integrate the delivery of digital content tightly into the database and (more programmatically) to increase the amount of digital material provided online. Much digital content had previously been stored outside of the Archon system and was not linked to archival descriptive records. We resolved to begin an aggressive program for augmenting descriptive records with digital content and took steps to increase the number of born-digital materials that we provided online. In particular, we began adding more photos to our Archon Digital Library. We developed a new section of our website (a pilot E-Records Repository³⁹) and added dual-direction links between the digital content and the associated descriptive records in the Archon database.

Based on the results of the goal analysis described under question 4 (p. 175), we also redesigned our goal conversion funnels to more accurately measure email conversions, to track downloads of digitized records/photographs or born-digital records, and to test the success of a pilot program to sell high-resolution copies of historical photographs.⁴⁰ Not only would such data show whether user needs had been met (albeit, in a very limited sense), it would also allow us to report specific data regarding online use (information of interest to administrators). More importantly, we completed further analysis of user progress along the goal conversion funnels to help us undertake website improvements to encourage even more digital content use and to drive decisions regarding future digital services. For example, we decided to track how often users searched for images, viewed digital content records, began an order process, and submitted an order. We hoped this would lead to increased use of the images, and possibly, sales of our high-resolution images, thus returning a modest supplement to the archives discretionary budget.

³⁹ The repository was launched as a proof of concept project using a simple PHP script to browse nested content on a Web server. See "University of Illinois Archives E-Record Repository," <http://www.library.illinois.edu/archives/Electronic%20Records>, accessed 1 October 2010.

⁴⁰ To measure the precise number of times particular pieces of digital content (such as images or PDF files in our digital repository) were being downloaded, we added tracking code to download links generated by our Archon database. The process of tracking downloads is described at Google Analytics, "Help," <http://www.google.com/support/googleanalytics/bin/answer.py?hl=en&answer=55529>, accessed 1 October 2010, and involves adding javascript on click events to each HTML anchor link that you wish to track. We configured the links so that the analytics report generated a "hit" each time the link was clicked, sorting all hits into a directory structure. This maximized the amount of information collected so that we could tell how often particular types of files and indeed, individual files, were being downloaded. For example, the tracking code for a link to download a digital image stored in our Archon database is encoded as `onclick="javascript: pageTracker._trackPageview('/digcontent/image/DigitalContentID=3956/fileID=3902')"`. By drilling the analytics reports, we can tell how many pieces of total digital content have been downloaded, how many of the total were images (as opposed to other content types such as AV files or PDFs), and even how many times a particular file was downloaded. Thus, we can easily see not only what general types of content are the most popular, but which particular images or files are most often downloaded over time.

Implementing a Continuous Improvement Program

Changes to the design of the website and the configuration of the new goals were completed by the end of August 2008. To compare website use before and after the changes, we ran the analytics reports for the comparable period (July 2009) and for a new baseline tracking period (July 2008 through June 2009), providing a basis for future longitudinal analysis. These key website performance indicators are shown in Table 1.

In general, the results show that the changes had desirable effects. Fewer users were bouncing from our site after viewing one page, and the average time that each user spent on the site had noticeably increased. Most importantly, email contacts had increased sharply, and this increase also corresponded with increased reference use of the archives. While the archives registered 734 physical analog consultations of our materials in July 2007, we registered 1,146 such transactions in July 2009, an increase of 56%.⁴¹ While correlation is not the same thing as causation, it appears likely that improved website utility facilitated more in-person and remote reference. In July 2007, 49 users viewed the email form, but did not necessarily email the archives, while in July 2009, 62 users actually completed and submitted the email form.

Though heartening, these results were even more striking given another salient fact: Not only did usage of our website (i.e., the number of page views) fail to increase between July 2007 and July 2009, it actually declined by nearly 11%. After drilling into the results, we discovered that the decline was entirely due to the fact that Google was referring far fewer people (in both absolute and relative terms) to our website. While more visitors were entering the site directly or by hard links from other sites (such as *Wikipedia*), the number referred by Google dropped by 47% from July 2007 to July 2009.

Careful analysis of the analytics reports placed us in a position to understand why the decline took place. Comparable pages generated by our Archon database output ranked much lower in Google's search results than they did in 2007, if they appeared at all. Cracking Google's ranking algorithms is notoriously hard,⁴² but it was apparent that something had affected the manner in which Google ranked our pages, even though we made no changes to the URL or to the structural elements in the page's HTML code. By installing the Google toolbar (which provides a 1 to 10 numerical ranking for how important Google thinks a page to be), we discovered that Google was indexing the top levels of our database but did not seem to be indexing lower levels, such as the actual

⁴¹ This increase was not an aberration but represents a sustained increase. During fiscal year 2008 (July 2007–June 2008), we experienced 8,024 uses; for the period from July 2009 through June 2010, we experienced 12,841 uses.

⁴² For example, see Ian Rogers, "The Google Pagerank Algorithm and How It Works," Vitaly-Friedman, <http://www.alvit.de/vf/en/web-development-the-google-pagerank-algorithm-and-how-it-works.html>, accessed 17 May 2009.

Table 1. Key Performance Indicators, University of Illinois Archives Website

	July 2007	July 2009	July 2009– June 2010
Visitors			
Unique visitors	12,008	7,882	101,674
Average time on site	01:56	02:38	3:11
Bounce Rate*	64.94%	59.09%	57.84%
Traffic Sources			
Referred by search engines	10,673 (75.9%)	6,252 (60.8%)	86,554 (57.35%)
Referred by Google	9,979 (70.9%)	5,255 (51.06%)	74,262 (49.21%)
Visits beginning in database	55.33%	64.29%	59.45%
Visits beginning on homepage	11.90%	16.06%	16.18%
Content			
Total page views	45,515	40,592	664,548
Database page views	29,375 (64.5%)	29,951 (73.8%)	456,945 (68.89%)
Goals Reached (conversion %)			
Emails after viewing series descriptions §	49 (0.35%)	62 (0.6%)	626 (0.41%)
Orders hi-res image after search	n/a	0 (0%)	34 (0.02%)
Downloads full finding aid ¶	697/11,164 (6.2%)	893/5,614 (15.9%)	15,240 (15.58%)
Downloads digital content	n/a	576/5,460 (10.5%)	10,426 (13.55%)
Reference Statistics			
Analog transactions	734	1,146	12,841
<p>* Percentage of visitors who leave site after viewing only one page.</p> <p>§ The figure for July 2007 overstates the number who actually emailed the archives since the measurement taken at the time represented people who reached the email form, not those who actually sent an email.</p> <p>¶ Manually calculated by dividing unique page views for downloaded resource by unique page views of its parent page.</p>			

descriptive records. For example, one of the most significant resources we hold is the Avery Brundage Collection, which is heavily used by sports historians. However, a search for the quoted phrase “Avery Brundage Collection” did not include a link to our database record describing the collection in Google’s first 100 results—even though a few of the top ten Google results link to that very page. Since the page includes an appropriate page title: “Avery Brundage Collection: University of Illinois Archives,” we looked to other factors that might be causing the problem, and it appeared that we had unwittingly violated at least two of the rules that Google provides for ensuring that pages are appropriately

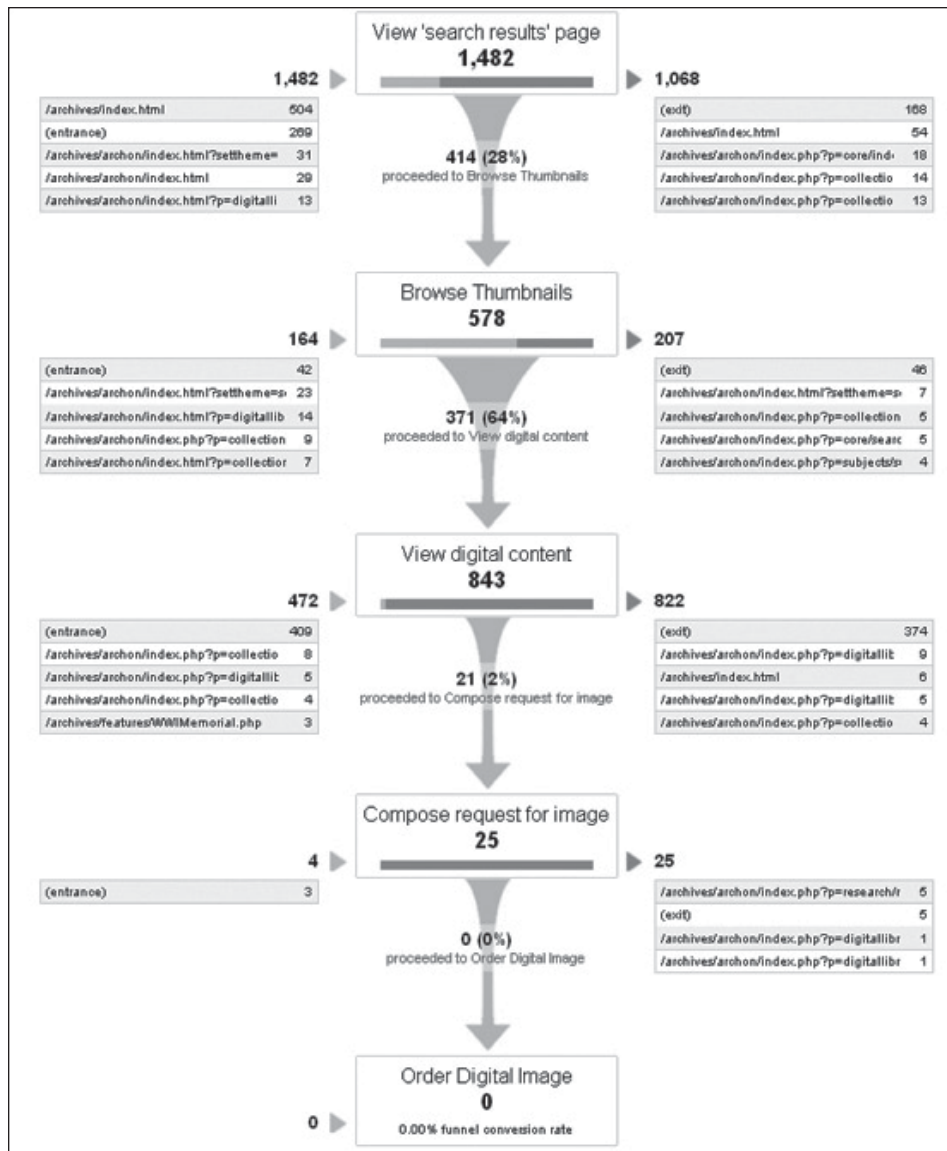


FIGURE 8. Goal conversion funnel for digital image viewing and ordering, July 2009.

indexed. We did not make use of the “description” meta tag and we did not use Google-friendly URLs.⁴³ Fixing the first problem was very easy and was accomplished in under ten minutes. Fixing the second problem is more complex, and we are investigating and implementing methods to ensure that our series-level

⁴³ “Google’s Search Engine Optimization Guide,” Version 1.1 (November 2008), <http://www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf>, accessed 17 May 2010.

descriptive records stand the maximum chance of being fully indexed by Google and other search engines.⁴⁴

Information provided by the analytics reports will help us drive programmatic decisions for a continuous improvement program regarding archival operations. For example, it will help us determine which content should be provided on the website and will also help us optimally structure the means of making that content available.

As noted above, we decided in 2007 to substantially increase the amount of digital content that we provide online. To the greatest extent possible using readily available technologies, we linked descriptive records and digital content held in our new e-records repository, which used a separate piece of software. We also began tracking the success of a new program to sell high-resolution copies of previously scanned digital images. Based on analytics data collected after the changes, we can see that some of these changes were successful in enhancing end-user access to our descriptive records and to digital content. For example, the finding aids were downloaded at nearly three times the previous rate, comparing July 2007 to July 2009, and at about twice the previous rate, comparing July 2007 to fiscal year 2010. Similarly, access to digital content increased, and about 10% of all page views of a descriptive record for digital content resulted in eventual download of the attached record.

However, not all of our changes were successful, and the analytics reports provide data that will justify and shape future decisions to expend scarce resources. Two examples illustrate the capacity of analytics to shape better end-user services.

First, our plan to begin selling high-resolution copies of images directly from the user's Web browser can only be described as an abject failure. In July 2007, not a single user completed the process of submitting an order form, in spite of the fact that our thumbnails page was viewed 578 times and the digital content pages for photographs 843 times.⁴⁵ The goal conversion funnel that we defined, shown in Figure 8, illustrates the pinch point: Only 25 users clicked the link to load the request form, and none of the 25 actually submitted this form. This problem persisted throughout the FY 2010 reporting period, as Table 1 illustrates.

This finding seriously calls into question our proposed plan to add e-commerce features (i.e., a shopping basket and credit card payment facility) to our site. To make a rational decision as to whether we should proceed, we will first revise the page layout, shown in Figure 9, as well as the order form.

⁴⁴ "Dynamic URLs vs. Static URLs," *Google Webmaster Central Blog*, <http://googlewebmastercentral.blogspot.com/2008/09/dynamic-urls-vs-static-urls.html>, accessed 1 October 2010.

⁴⁵ For FY 2010, the request form was submitted only 35 times, a goal conversion rate of only 0.03% against all digital image search requests submitted by users.

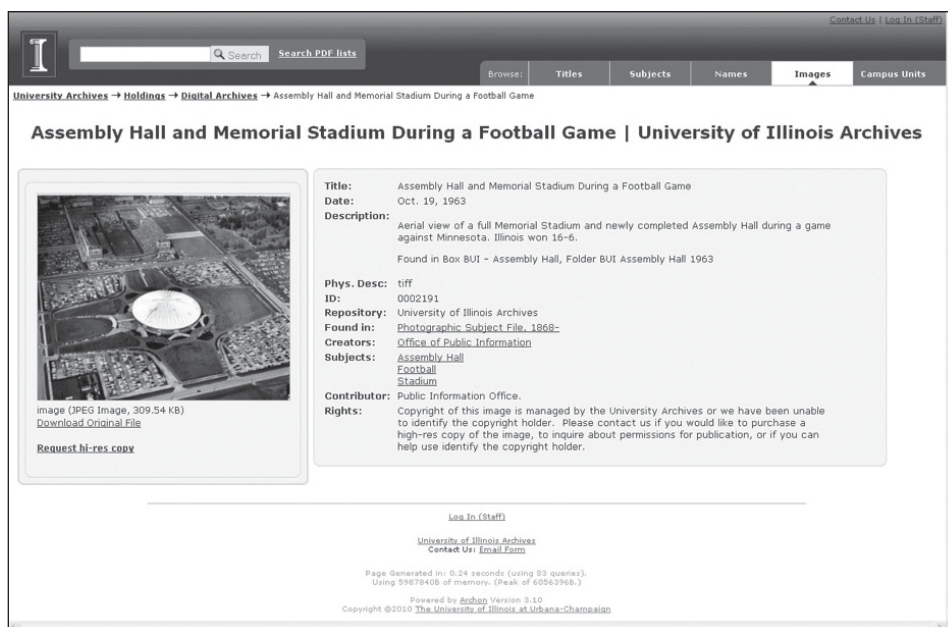


FIGURE 9. Digital content page design (to be redesigned).

How should we revise that page? Even though the form is clearly linked from the digital object context record and self populates, so few users clicked the link to load the form that we suspect it is either misleading or not prominent enough. Specifically, use of the analytics software helped us determine that the link “download original file” above the “request hi-res copy” link is inaccurate, since the original file is not actually provided, but only a slightly larger version of the medium-sized image. The link will be changed to read simply “download,” and the “request hi-res copy” will be revised to read “order hi-res copy,” which is more accurate. If these changes succeed in reducing the percentage of users who drop from the funnel, we will then be able to gauge the true level of interest in ordering high-resolution copies of images. In other words, we will be able to decide whether we would be justified in using resources to develop an e-commerce system to automate the process of ordering, receiving payment, and delivering images to our users. Such a decision will need to be strongly influenced by other work that suggests that the market for such historical imagery is fairly small.⁴⁶

Second, our delivery method for textual electronic records, such as PDF or Word documents, similarly failed. As noted above, an E-Records Repository was

⁴⁶ See Ricky Erway’s description of RLG’s attempts to license materials from its Cultural Heritage Initiative, “Seeking Sustainability” (2008), <http://www.oclc.org/programs/publications/reports/2008-03.pdf>, accessed 1 October 2010.

UIUC Archives → Electronic Records → University Archives right-pointing arrow

University of Illinois Archives E-Records Repository			
File	Description	Size	Modified
.. (parent folder)			
0101802 Trustees Reports	Board of Trustees Reports (Record Series 1/1/802)	[dir]	2008-Jul-07
0402817 Senate Committee Library	Senate Committee on the Library Reports (Record Series 4/2/817)	[dir]	2008-Jul-11
1105832 Transportation Engineering Studies	Illinois Cooperative Highway & Transportation Series (Record Series 11/5/832)	[dir]	2008-Jul-09
1208049 Champaign Greenways Trails	Champaign County Greenways and Trails Committee Publications (Record Series 12/8/49)	[dir]	2008-Jul-09
1505046 Gregorio Weber Papers	Gregorio Weber Papers (Record Series 15/5/46)	[dir]	2008-Jul-14
1535057 Boris Balinsky Memoir	Boris Balinsky Memoir (Record Series 15/35/57)	[dir]	2008-Jul-09
2503801 Registers Course Catalogs	Catalogs and Registers (Record Series 25/3/801)	[dir]	2008-Jul-09
2503810 Student Enrollment Statistics	Student Enrollment and IPEDS Data (Record Series 25/3/810)	[dir]	2008-Jul-14
2507801 Student Code	Student Code (Record Series 25/7/801)	[dir]	2008-Jul-09
2801806 Stadium Renovation Reports	Stadium Renovation Reports and Publications (Record Series 28/1/806)	[dir]	2008-Jul-09
3501801 Library Annual Reports	Library Annual Reports (Record Series 35/1/801)	[dir]	2008-Jul-10
3501826 Library Friendscript	Library Friendscript (Record Series 35/1/826)	[dir]	2008-Jul-09
6007801 UIC Library Annual Reports		[dir]	2009-Apr-23
0 Files · 13 Folders		Total size: 0.0	

FIGURE 10. E-Records repository interface (auto index script).

implemented as a quick way to make such records browsable. I installed and configured an open-source script called Auto Index to provide a simple browsing mechanism for files that lack item-level metadata. The interface is shown in Figure 10.

During July 2009, records from this repository were downloaded only 35 times; for fiscal year 2010, 2,151 times. Image files, which are managed in our Archon database (and thus are fully searchable and viewable in the same interface that manages descriptive information for non-digital materials), were downloaded 584 and 7,607 times, respectively, over the same periods.

Why were so few textual documents downloaded? It is possible that the records are simply too prosaic and thus of less research interest than images, but this is unlikely since the files include core university records that are heavily used (such as trustees' proceedings, enrollment statistics and the *Student Code*), as well as the records of external organizations, such as the American Library Association and the American Association of Law Libraries, which are in heavy demand. It seems more likely that users are simply not finding the records. Upon further analysis, we determined that while many of the records are linked to the database, the digital content is, at best, three or four clicks removed from the descriptive metadata. In retrospect, the decision, made in 2007, to remove e-records storage from the Archon database, was wrong. We should have expended the time and money to integrate their access fully with the database, as we did for images.

In the best of all worlds, the next version of our Archon database software would address this issue. However, project development for the Archon project is currently frozen, pending the results of the Archon/Archivist Toolkit planning process.⁴⁷ Nevertheless, this finding should be taken into consideration

⁴⁷ Susan Harum, "ArchivesSpace Planning Project," post to Archivist Toolkit User Group Listserv, 14 May 2010, <https://mailman.ucsd.edu/pipermail/atug-l/2010/002437.html>, accessed 1 October 2010.

during the specification development process for the proposed software. Generally speaking, the profession lacks information about the best methods to facilitate access to electronic records, and these limited findings suggest that divorcing object storage and access from object description is correlated to lower rates of access. Any project to link disparate systems will be less than fully successful unless significant attention is paid to facilitating effective bidirectional access to metadata and its associated digital objects.⁴⁸

Web analytics software provides one method to understand user preferences and behavior, and, more importantly, to improve archival services. These examples are intended to illustrate how the software can be used. Repositories cannot expect to simply install the tracking code, walk away, and dutifully report statistics to administrators. Repositories must clearly understand website goals, select and configure appropriate analytics tools, and expend a moderate amount of time and effort to interpret the reported results. There is considerable scope for experimentation. For example, other repositories may wish to implement site search, event tracking, and site overlay (each of these concepts is explained fully on the Google Analytics help pages). Or a repository might launch its analytics program to complement qualitative analysis of Web users' experiences; we certainly hope to pursue this option at the University of Illinois Archives.⁴⁹ In either case, Web analytics gives archivists the power to design and implement an effective Web delivery program—if we use it wisely.

⁴⁸ The *Archon/AT Integration Project Hi-Level Functional Requirements* developed as part of the ArchivesSpace planning activities indicate that object storage will be not be included in the integrated application. As a result, images and other objects currently stored in Archon instances will need to be migrated to other applications. However, the findings of this study were not available at the time the *Hi-Level Requirements* were drafted, and there may be scope to reconsider how object storage and linking best be facilitated in the proposed application, <http://www.archivesspace.org/documents/AT%20Archon%20HiLevel%20Reqs-FINAL-2010127.pdf>, accessed 1 October 2010.

⁴⁹ The Archival Metrics Project provides appropriate survey mechanisms for online finding aids and websites, "Archival Metrics Toolkits," <http://archivalmetrics.org/node/10>, accessed 1 October 2010. See Wendy M. Duff, Elizabeth Yakel, Helen R. Tibbo, Joan M. Cherry, Aprille McKay, Magia G. Krause, and Rebecca Sheffield, "The Development, Testing, and Evaluation of the Archival Metrics Toolkits," *American Archivist* 73 (Fall/Winter 2010): 569–99.