

# Measuring audience attention across multiple channels for a new Web site

Joe Pagano  
Library of Congress  
101 Independence Ave., SE  
Washington, DC 20540  
1 (202) 707-2488 | jpag@loc.gov

*[June 14, 2007] The views in this paper are the author's only, and are not an endorsement of any product or company, and do not necessarily reflect the views of the Library of Congress or the federal government. Data presented in this paper should not be considered official Library of Congress statistics.*

## ABSTRACT

This paper presents analysis done on traffic data for a new Web site section offered by the Library of Congress. The period analyzed covers the first two weeks of the site being publicly available. Large cultural Web sites, by virtue of their mission, often serve diverse groups of users and continually add to their Web site content. The additions go beyond routine maintenance and updating of pages and include the introduction of entirely new products that complement existing offerings. New site sections may have new features targeted to new audiences. While it may be easy to simply roll a new site into existing metrics reports, and only summarize the increased traffic to the “whole” site, valuable information can be gained by developing customized analysis for new sites, taking into consideration the unique features of the site and its intended audience focus. This paper describes a customized analysis done for one particular site, the *Chronicling America* site, introduced in March 2007. The analysis reviews the Web activities, most employed by users, to drive user attention to the new Web site.

## Categories and Subject Descriptors

C.4 PERFORMANCE OF SYSTEMS - *measurement techniques, performance attributes*

## General Terms

measurement, human factors, metrics, analytics, audience analysis

## Keywords

analytics, metrics, measurements, log files

## 1. INTRODUCTION

The Library of Congress Web site serves a worldwide audience, with the priority audiences being comprised of the Congress and the American public. The Library divides its public audience into several large groups including: the educational community (students, teachers), librarians, publishers, scholars/researchers and the general public, including families and children. Since

the start of 2007 the Library introduced several major new sections of its ever-growing site. They have included heritage month sites for African Americans, Asian Americans, and Jewish Americans, an online exhibit on the MacDowell Colony, and lastly *Chronicling America*, an archive of historic American newspapers. Each of the sites have unique aspects and goals, that require differing analytic strategies for the Library to maximize what is learned so future sites can be introduced in more successful and meaningful ways. Important factors to be considered during the introduction of a new site are: whether the site is discoverable (i.e. will the correct search terms lead users to the site) and where the site benefits one unique audience more than others, whether a specified audience is aware of the site's existence. The two factors are somewhat related. We often think of search in terms of answering an immediate question, but not every member of an audience searches daily for life long topics of interest. In these cases, questions and searches may not be immediate and finite, but rather users develop a set of tools they can rely on over a long period of time. To that extent, search and discovery take on aspects of developing a personal research library. This different perception makes online word of mouth (OWOM) and referrals more important than in information seeking situations where the answer has a value based on a discrete time constraint. There are more ways to find relevant information on the Internet than ever before, and savvy online users are incorporating new behaviors and tools into their search and discovery behavior. This is probably in part to compensate for the shortcomings of the traditional search engines as algorithms are not advancing quickly enough to be able to present the most useful topics to users on a consistent basis, especially where the desired information is more associated with research over an extended time period on a consistent topic.

## 2. BACKGROUND

In March 2007, the Library of Congress introduced the “*Chronicling America*.” Web site. The site is a beta site, providing access to select U.S. newspapers. The National Digital Newspaper Program (NDNP), a partnership between the National Endowment for the Humanities (NEH) and the Library of Congress (LC), is responsible for overseeing the project. The NDNP long-term goal is to develop a large Web-based, searchable database of historic U.S. newspapers, representing many cities and states. With support by NEH, this rich digital resource is being developed and maintained at the Library of Congress.

Chronicling America was soft launched on March 14, 2007, and formally launched on March 21<sup>st</sup>, with a home page link and press release (<http://www.loc.gov/today/pr/2007/07-061.html>). Referrer and general metrics data were collected for the period from the soft launch date through a week after the formal launch date. The data were analyzed for various parameters, with the intent of eventually comparing how awareness of this and future Library of Congress sites spread on the Internet. The goal is to better understand this process, so the Library can more efficiently and successfully launch future sites, using a methodology that serves users who will benefit the most from the content.

### 3. METHODOLOGY

Referrer and visit data were pulled for a two week period starting March 14<sup>th</sup>, the date the site was first available outside of the Library’s firewall. The log data were imported into a database, and various text extractions and calculations were performed on the URL data to scrutinize referring sources. The primary technical process used in this study was text identification and summation, based on the assumptions noted here. It was assumed that any domain, sub domain, or path that included the text “mail” was from an online email service. A similar assumption was made in the case of genealogy, where the assumption was based on the text “geneal” and after reviewing several of sites generating large amounts of traffic, was expanded to include “researchguides” and “findagrave”, because these text snippets were in the domains of sites that appeared to be primarily devoted to genealogy. The database summed up instances of referrals for each text snippet mentioned, along with the number of visits associated with each instance of a referral. The charting analysis was then done using the visit data. In addition to analyzing “categories” of sites, these specific domains were included: del.icio.us, Google, Yahoo, the University of California at Berkeley, a project partner, and the Library of Congress (note: referrer traffic from the Library’s main site includes internal along with public traffic). Using assumptions such as the ones noted above, the data was analyzed for these categories: email, blogs, genealogy, groups, and search. The categories are not mutually exclusive and a site could be counted, as being both from Google and “groups”, and as both a “genealogy” and a “blogging” site. After the appropriate calculation fields were created and processed, the visits data were graphed, to show how each site or category compared when driving visits to the Chronicling America Web site.

## 4. ANALYSIS AND RESULTS

### 4.1 Soft versus hard launch

The first and easiest comparison was between soft launch and hard launch data. The Library provided access to the University of California at Berkeley, as well as other partner institutions, during development and testing, but site access was limited based on institutional domains. After the soft launch date, anyone with the correct URL was able to view the Web site. During the week of the soft launch, the site was demonstrated to at least one educational group visiting the Library. Traffic to the site during the soft launch period indicates that users who were aware of the site did not generally publicize it much until the official launch date. The exception is in the case of at least one user’s apparent interest in sharing the URL with genealogists,

since a noticeable “bump” (Fig. 1) on March 15<sup>th</sup> is derived from genealogy site traffic.

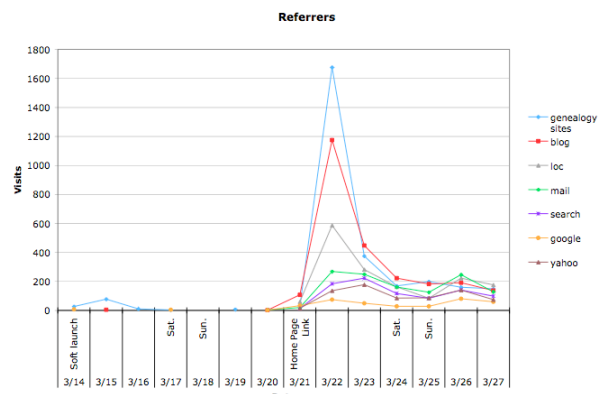


Fig. 1

### 4.2 Referrers versus typed / bookmarked

Both referred visits and typed / bookmarked (TBM) visits show a similar pattern for the two week period, except for a generous gap -up for referred visits on the day after introduction. Over the next several days the gap decreases and eventually equals TBM (Fig. 2). Based on TBM strength it can be concluded that many people view this site as a resource worth remembering. It will be interesting to see how this trend develops over an extended period of time.

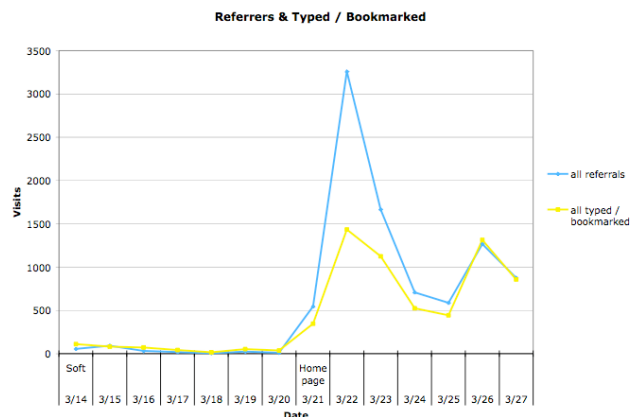


Fig. 2

### 4.3 Expanding attention

The next segment of analysis was for the week after formal introduction of the site. The major drivers of traffic to the Chronicling America Site, were in descending order: genealogy sites, blogs, referrals from the Library’s site (including the home page), email, and lastly search. The fact that blogging and email both ranked above search (Fig. 3) indicates the important role

“online word of mouth” (OLWM) can now play in increasing the number of users showing attention to a Web property.

worthwhile to learn more about the process of new site introduction, so that institutions can do it more efficiently and successfully. Given that “online word of mouth” can be a very successful way to drive traffic to a site, blogs and email should play an important role in any site introduction.

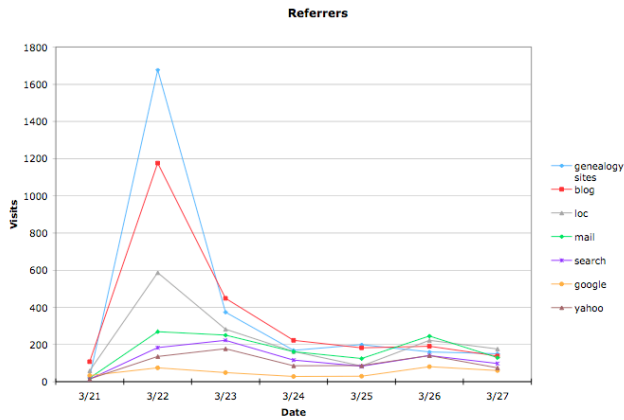


Fig. 3

#### 4.4 Search versus online word of mouth

Looking closely at the five-day period from March 23rd to the 27<sup>th</sup>, a new trend can slowly be seen developing. Referrers from blogs and email show a sharp decline, while search keeps level or may, in fact, be increasing. The peak on the 26<sup>th</sup> for email (Fig. 4) indicates that, at least initially, OWOM functions slower in email, than in blogging, which would make sense given that non-spam email is often exchanged between individuals, blogging is a one to many relationship all of the time. It is expected that there will be increased referrals from search engines, as increased links to Chronicling America, help to increase the site’s ranking on search engines.

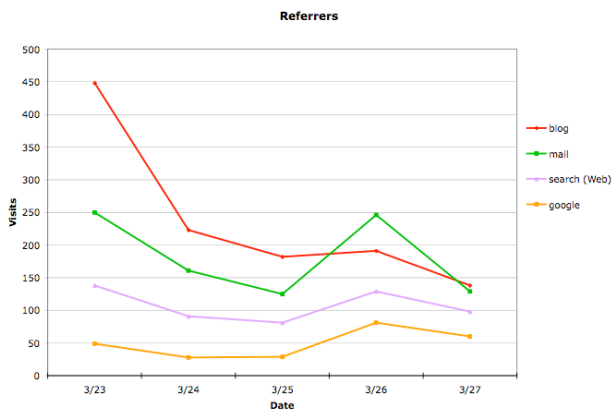


Fig. 4

### 5. CONCLUSION

A Web site will display different patterns of attention origination during the introduction period. These patterns reflect different characteristics of content and audience. In the case of Chronicling America, OWOM played a more significant role than search in initially focusing user attention on the site. It is

---

This paper is authored by an employee of the United States Government and is in the public domain. CAMA'07, June 23, 2007, Vancouver, British Columbia, Canada.