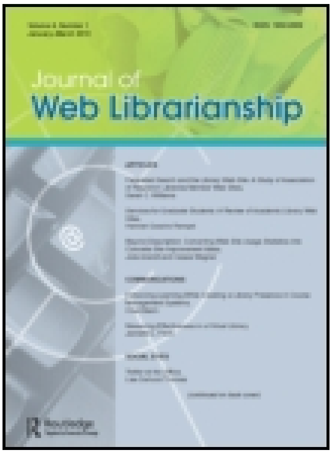

Journal of Web Librarianship
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/wjwl20

Using Web Analytics to Evaluate the Effectiveness of Online Maps for Community Outreach

Howard Veregin[a] & A. J. Wortley[a]
[a] Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin, USA
Published online: 28 Feb 2014.

To cite this article: Howard Veregin & A. J. Wortley (2014) Using Web Analytics to Evaluate the Effectiveness of Online Maps for Community Outreach, Journal of Web Librarianship, 8:2, 125-146, DOI: 10.1080/19322909.2014.873659

To link to this article: http://dx.doi.org/10.1080/19322909.2014.873659

Journal of Web Librarianship, 8:125–146, 2014
Published with license by Taylor & Francis
ISSN: 1932-2909 print / 1932-2917 online
DOI: 10.1080/19322909.2014.873659

Using Web Analytics to Evaluate the Effectiveness of Online Maps for Community Outreach

HOWARD VEREGIN and A. J. WORTLEY
Department of Geography, University of Wisconsin-Madison, Madison, Wisconsin, USA

This study uses Web analytics to evaluate the effectiveness of online maps to support community outreach efforts. The Wisconsin State Cartographer's Office has long used Web technology, including online map applications, to deliver geospatial information to citizens of Wisconsin. This study focuses on one of these applications, the Wisconsin Historic Aerial Image Finder (WHAIFinder), which provides online access to a rare collection of aerial photographs from the 1930s. The goal of this study is to assess whether WHAIFinder has succeeded in providing broad statewide access to the digital collection. As such, the study asks whether usage rates for the application are essentially uniform across the state or if there are regions where usage is lower or higher than expected. The study's approach is inherently geographic due to its concern with patterns of use reflecting local variations in demand, access, and knowledge. Regression techniques and geographic analysis methods were used to model relationships in the data, visualize statistics using maps, and analyze geographic patterns. Results show an overall pattern of uniformity but also reveal zones where usage deviates significantly from expected values. This study's methods are potentially useful to researchers in other fields interested in evaluating geographic patterns for online applications intended for outreach and information delivery.

KEYWORDS *Web analytics, community outreach, service effectiveness, Web maps, geospatial technology*

Received 13 September 2013; accepted 13 October 2013.

Address correspondence to Howard Veregin, Wisconsin State Cartographer, Department of Geography, University of Wisconsin-Madison, 550 N. Park St., Madison, WI 53706. E-mail: veregin@wisc.edu

INTRODUCTION

This study employs Web analytics methods to evaluate an online application that delivers digital geographic data to the public. The Wisconsin Historic Aerial Image Finder (WHAIFinder, http://www.sco.wisc.edu/whaifinder/whaifinder.html) was developed to provide online access to a rare collection of historic aerial photographs of Wisconsin (Veregin et al. 2010). WHAIFinder allows users to discover and download these photographs remotely from any computer connected to the Internet. The application expands utilization of the photograph collection, especially for users who live far from the physical collection. WHAIFinder also accommodates a wide range of user expertise, from the general public to geospatial professionals and university researchers. The overall objective of the WHAIFinder project was to design and deploy an online application providing data access for all Wisconsin residents and in all regions of the state.

WHAIFinder was developed by the Wisconsin State Cartographer's Office (SCO), the Robinson Map Library (Department of Geography, University of Wisconsin-Madison), and the University of Wisconsin Digital Collections Center. WHAIFinder is one of several online applications that support the SCO's mission of fostering the equitable exchange of geospatial information to deliver services to the citizens of Wisconsin (http://www.sco.wisc.edu/sco/aboutsco.html). The SCO's specific responsibilities, as described in state statute, focus on information dissemination to professionals and the public (State of Wisconsin 2012). The Internet has long played an important role in these efforts (Veregin and Kennedy 2012).

The main goal of this study is to evaluate the effectiveness of the WHAIFinder application. Here, the term "effectiveness" means the degree to which WHAIFinder has succeeded in providing broad statewide access to the digital photograph collection. To evaluate effectiveness, the study used regression analysis to quantify the relationship between WHAIFinder usage (the number of users accessing the application over time) and population estimates in Wisconsin communities derived from the U.S. Census. Some of the specific questions examined were the following: Is there a direct, positive relationship between WHAIFinder usage and population in Wisconsin? If so, what are the statistical parameters defining this relationship, and what is the strength of the relationship? Are there regions of the state where usage rates are lower or higher than expected relative to population? Is usage concentrated in certain regions of the state? Does access to broadband Internet affect these relationships?

In contrast to many other studies of Web usage, this study's perspective is strongly geographic. The authors are interested in geographic patterns of usage and the relationship of these patterns to underlying population distributions. To this end, in this study some conventional tools and methods of geography and geographic information systems (GIS) including geocoding (locating observations geographically), visualization of statistics using maps, and statistical analysis of geographic patterns were adopted.

Online applications similar to WHAIFinder have been deployed in several states to provide broader access to physical collections maintained by libraries. Two examples include the University of Florida's "Aerial Photography: Florida" site (http://ufdc.uflib.ufl.edu/ufdc/?c=flap) and the University of Minnesota's "Historical Aerial Photographs Online" site (http://map.lib.umn.edu/mhapo). Such examples indicate a growing number of geospatial applications adapted to online data discovery and access in libraries (National Research Council 1999; Goodchild 2004; Bodenhamer, Corrigan, and Harris 2010; Goodchild and Janelle 2010). Given this growth, the authors anticipate that some of the concepts and methods described in this study could be employed by other libraries to assess the effectiveness of their own online applications.

Looking beyond the immediate goals of this project, the ultimate aim of this line of research is to gain insights into return on investment (ROI), which is the ratio of benefits derived from an activity to the cost and effort associated with developing and maintaining that activity. ROI assessments have been important in many sectors for years, and their use in libraries has been rising recently due to a decrease in funding and a concomitant rise in demand for accountability and improvements in efficiency. The growing field of library valuation allows researchers to assess the monetary value of library programs and services and show efficient use of tax dollars in cost-benefit terminology (Missingham 2005; Aabø 2009). Increasing sophistication in this field has led to efforts to measure secondary impacts of library services on the rest of the economy, account for the intangible social dividends of libraries, and demonstrate the value of libraries to managers, stakeholders, clients, and society as a whole (Missingham 2005; Aabø 2009). Quantification of usage rates over time and space, as exemplified in this WHAIFinder study, is a necessary first step in such ROI assessments.

LITERATURE REVIEW

Web analytics is defined as the "measurement, collection, analysis and reporting of Internet data for the purposes of understanding and optimizing Web usage" (Web Analytics Association 2008). The real value of Web analytics lies in the discovery of business opportunities by studying user habits and behaviors (LaPointe 2010). For the purpose of this study, Web analytics will be used to identify opportunities for improving performance on specific program goals, specifically those related to the SCO's information dissemination and community outreach mission. The authors' ultimate aim with this study is to improve information delivery effectiveness by identifying factors

TABLE 1 Web Performance Metric Model (after Wood et al. 2003)

	Analysis method			
	1. Usability testing	2. User feedback	3. Usage data	4. Technical performance data
Description	Formal assessments based on user testing in a controlled environment	Collection and analysis of written or verbal feedback from users	Collection and analysis of quantitative data about Web site traffic	Measurement of Web site and Internet performance
Supported functions	Enhance Web site design, navigation, functionality, and topology to improve user experience	Assess user profile, demographics, satisfaction, and use	Evaluate pages viewed, number of visits, unique visitors, searches run, etc.	Quantify latency, availability, data transfer rates, etc.

contributing to more equitable and engaged use of the SCO's online mapping applications throughout the state.

While Web analytics originally had a strong e-commerce focus, its methods have been adapted for use in many other areas. Even nonprofit Web sites exhibit underlying indicators of engagement that can be important for measuring performance and effectiveness (Turner 2010; Marek 2011; Durante and Wang 2012). Using Web analytics effectively requires attention to key performance indicators (KPIs) that precisely reflect the specific objectives of the analysis.

According to a study by Wood and colleagues (2003), Web analytics methods can be classified based on the functions they support (Table 1). In the Wood et al. study, these functions were shown to be appropriate at different phases of a Web site's life cycle, including initial design and development, ongoing maintenance efforts, and implementation of enhancements. The classification can be used to select specific approaches and metrics relevant to a particular situation.

The present study focuses on the usage data method; the other methods are not directly relevant to the questions being addressed. Since the WHAIFinder application has already been developed, modifying the user interface based on usability testing is not a top priority at this time. Likewise, the user feedback method is not directly applicable to the issues being addressed because the focus is on service effectiveness rather than patron satisfaction. Performance data has some relevance to the study because the authors hypothesize that low-bandwidth users may experience problems with large WHAIFinder image files, but for the most part technical performance issues will not be addressed. Since the primary concern lies in assessing WHAIFinder effectiveness in terms of usage levels, this study's focus is the usage data method. An added dimension of the study is a focus on locational data to visualize and analyze geographic patterns of usage across Wisconsin.

Several different statistics are available to measure Web site usage, including pageviews, visits, and visitors. A pageview is calculated every time a page on a Web site loads, but pageview statistics can significantly overestimate actual usage since multiple pageviews can be generated in a single session. A visit or session is a period of interaction between a Web browser and a Web site. A visitor is essentially synonymous with a unique user. This study focuses on visitors as the key statistic due to the authors' interest in relating usage to local population data.

Libraries are at the forefront of using Web analytics for assessing Web technology's effectiveness in meeting information dissemination and outreach goals. In part, this is because Web sites have become an essential component of library outreach (Marek 2011). Web sites require specific measures of effectiveness that go beyond traditional measures of library effectiveness and emphasize patrons' needs and requirements (Turner 2010; Marek 2011). Since many libraries are operating under tighter budgetary constraints than in the past, part of this analytical focus reflects a need to justify expenditures from a financial perspective (Marek 2011).

In a study of Web analytics for academic libraries, Wei Fang (2007) evaluated the Rutgers-Newark Law Library Web site by tracking overall usage and assessing user behaviors to identify possible improvements to the site's design. Changes to the site were tested using metrics like time on site and the ratio of new to returning visitors, which revealed that the changes helped improve navigation and patron loyalty. Jannette Finch (2010) focused on measuring service quality for a small virtual academic library at the College of Charleston and used KPIs like the saturation rate of student population (market penetration), the ratio of sessions to courses offered, referred traffic sources, and visit duration. Kim Durante and Zheng Wang (2012) evaluated search and retrieval tools used by Emory University Libraries. They proposed an assessment framework based on KPIs that the library could use to monitor important dimensions of user interaction with the aim of prioritizing limited library resources and improving end user satisfaction and participation rates. These KPIs were constructed to develop a picture of the user population, measure changes to the size of the user base over time, compare the relative popularity of different online discovery services, and assess user satisfaction, user engagement levels, and task completion rates.

Similar types of analyses have been applied to Web sites supporting educational initiatives. For example, Casey Jones and colleagues (2004) reported on early efforts to collect Web usage data for the National Science Digital Library to assess how effective the site was in reaching the general population. Likewise, Emily Puckett Rodgers (2011) reported on an effort to evaluate the effectiveness of the Open.Michigan initiative and the value it adds to the global learning community. The goal of the evaluation effort was to strengthen accountability within Open.Michigan's home institution and across the open education movement. Quantifying resource use with Web analytics allowed researchers to determine how well organizational goals and publication methods matched the needs of the teaching and learning community, and understand which resources are most valuable. Furthermore, the evaluation provided a necessary baseline for ROI assessments based on overall investment levels in open educational resources and online delivery.

While academic libraries often have a primary mission to serve their home institutions, initiatives like Open.Michigan have a broader geographic focus and a resulting need to map and analyze visitors' locations. The Open.Michigan study found that a large percentage of its users (35 percent) were international, with just 21 percent of visitors coming from the vicinity of Ann Arbor, Michigan. This type of consideration is also relevant to this WHAIFinder study, given the authors' interest in assessing the uniformity of usage rates geographically across Wisconsin.

Web analytics has also been adopted for use in science communication. Oli Usher, Oana Sandu, and Lars Lindberg Christensen (2012) described efforts to evaluate the European Space Agency's outreach and educational programs for the Hubble Space Telescope. They included Web sites, videos, and social media in their analysis. Hao Tian and colleagues (2009) examined data for the chronic fatigue syndrome Web site of the Centers for Disease Control and Prevention to assess public utilization of the Web site. Of particular relevance to this study, the researchers analyzed visitor locations and computed visitor density for each state. Beatriz Plaza (2009) developed a simple repeatable time-series method to assess Web site usage for a small nonprofit Web site. One unique aspect of Plaza's study is its focus on the quality of users' visits, as measured by visit duration, not just by the quantity of these visits.

These studies have important implications for the current study. In the most general sense, they demonstrate the widespread use of Web analytics for evaluating online information delivery by libraries, content providers, and other organizations. Second, they demonstrate a diversity of approaches for Web site evaluation, from analysis of user interaction with specific Web pages to interpretation of usage data over time. Third, these studies demonstrate the validity of using visitor usage data as a metric for assessing Web site access and outreach effectiveness. Finally, these studies show that there is interest in mapping and analyzing users' geographic locations, especially when program goals include broad geographic dissemination of information.

Usage rates for an online application or Web site may be affected by an array of factors, some endogenous to the application or site itself, and some exogenous. At least some of these factors are expected to affect WHAIFinder usage rates and create visible geographic patterns in usage. To the degree that these patterns are non-uniform geographically, they will have some impact on the uniformity of WHAIFinder usage across the state.

There are significant differences among groups in terms of access to and knowledge of information technology, partly as a result of socioeconomic and educational factors. Research on the "digital divide" and inequalities in Internet access has focused on factors including education, race and ethnicity, gender, age, income, and rural location (Hargittai 2002). These factors not only affect physical access to the Internet but also skill and knowledge levels, use patterns and autonomy, and the availability of support networks for assistance. Even when physical access is not an issue, less experienced users may have more difficulty finding content online, using Web applications, and downloading and installing browser plugins or other software (Hargittai 2002). Geographically, usage might be expected to be higher in areas where the underlying demographic profile reflects greater knowledge and experience with the Internet, such as areas with higher educational attainment and lower-than-average median age.

Uneven usage rates for WHAIFinder may also be associated with variations in awareness and knowledge. There is certainly a link between increased awareness and increased usage; internal monitoring of SCO Web site traffic, for example, consistently shows spikes in visits following outreach efforts and media reports (such as conference presentations and news articles about SCO activities). Research also shows the importance of referred traffic like visits originating from an independent Web site through an embedded link (Plaza 2009). This suggests geographic clusters of high usage would occur in locations where the referring traffic originates, such as libraries and university campuses that are early adopters of the mapping tool.

The physical availability of broadband Internet may also affect usage rates, hence recent efforts to map broadband access nationwide via the National Broadband Map (http://www.broadbandmap.gov). The National Broadband Map is an attempt by the National Telecommunications and Information Administration and the Federal Communications Commission to map broadband infrastructure at a relatively fine geographic scale across the United States. The broadband map allows underserved areas to be identified and policies and programs at national and state levels to be developed to alleviate disparities in access (Neville 2013b). The National Broadband Map, along with NTIA reports and other studies, shows significant gaps in access across the country have continued, particularly along a rural-urban divide, but also within urban areas (Neville 2013a). These effects may be heightened for Web-based mapping applications because of their data-intensive nature (Kyem and Saku 2009). In short, broadband access has a strong geographic footprint, and the authors expect that disparities in access will be reflected in variations in WHAIFinder usage rates.

Usage may also be affected by the distance separating users from the hosting site, which can affect network performance and usability. It is commonly believed that geographic distance has no effect on Internet interaction levels, but empirical work challenges this idea. Yong Wang, Phillip Lai, and

Daniel Sui (2003), investigating the correlation between physical distance and network performance for US higher education Web sites, found distance to be positively correlated with latency (length of time for data to travel through the network). Latency has a negative effect on user satisfaction especially when large volumes of data must be transferred. WHAIFinder images can be quite large (approximately 30 Megabytes in size), hence distance-induced latency may be a factor contributing to reduced WHAIFinder usage in more remote parts of the state.

Some of these factors are more easily addressed than others. This study focuses on the infrastructure issue by accounting for access to broadband. The authors are also interested in how awareness of WHAIFinder might affect demand but have not yet developed a way to measure this relationship. Factors over which the study has no control (e.g., socioeconomic and educational) are necessarily a lower priority for the authors despite their significant effects on Internet participation rates.

DATA AND METHODS

The WHAIFinder application is a Web-based interactive mapping tool allowing users to identify, select, and download scanned historic aerial photographs of Wisconsin. The original photographs were acquired by the U.S. Department of Agriculture in the 1930s. This collection is the earliest systematic aerial survey of the state. It consists of over 38,000 photographs, physical copies of which reside in the Robinson Map Library at UW-Madison. High-resolution scans of the photographs, along with associated metadata, reside in a repository maintained by the University of Wisconsin Digital Collections Center built on the Fedora Commons platform.

WHAIFinder was launched in February 2011 and has become one of the SCO's most popular online applications. Eighty percent of traffic is from users in Wisconsin, with approximately 19 percent from other parts of the United States, and approximately 1 percent from other countries.

The WHAIFinder map interface (Figure 1) was designed to meet the needs of a diverse community. For novice users, simple navigation that mirrors other online mapping applications is available. More advanced functions include the ability to change basemaps, customize the interface by turning layers on and off, and interact with the photographs to assess change over time. Users can download the photographs in several different formats and resolution levels. For additional technical details on WHAIFinder, see Howard Veregin and colleagues (2010).

This study's source of Web usage data was Google Analytics (http://www.google.com/analytics), one of the most popular Web analytics tools. The tool was chosen for many of the same reasons other nonprofits choose it, including its advanced capabilities, free availability, and ease of use.

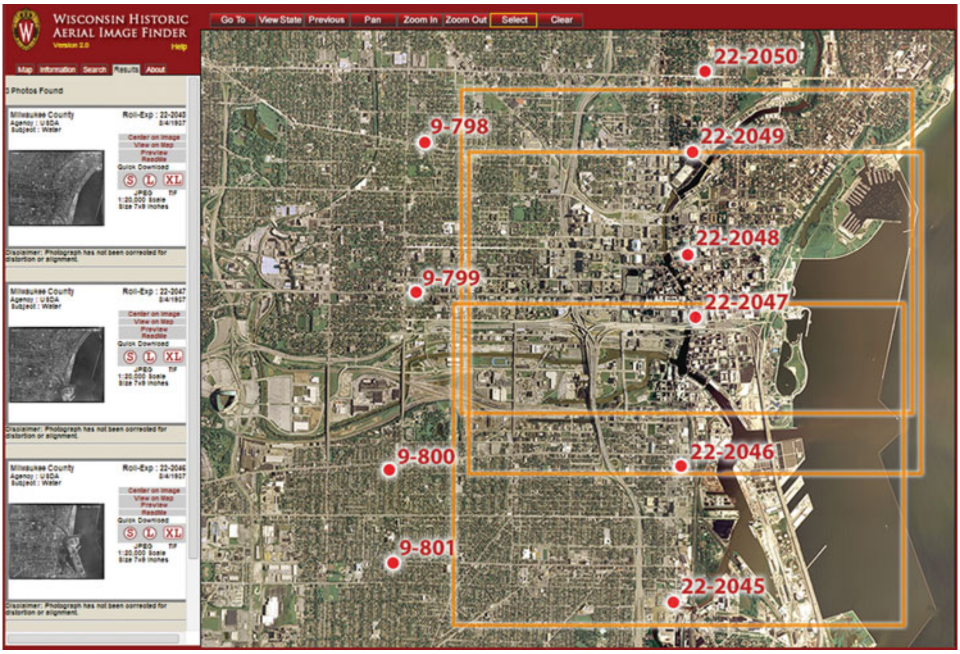

FIGURE 1 WHAIFinder interface. (Color figure available online.)

This study's KPI is the number of WHAIFinder visitors stratified by time and location. Usage was correlated with local population statistics; the number of visitors is the best surrogate for the actual number of users. To generate visitor totals, it was necessary to run a customized report in Google Analytics. The number of unique visitors for a date range of June 15, 2011 to June 14, 2012 was calculated with location as a secondary variable. Google provides statistics on visit location down to the city level; for WHAIFinder, Google identified almost 2,500 cities for the selected year-long date range. This study excluded cities outside of Wisconsin since the focus was on service delivery within the state.

Despite Google's "city" nomenclature, these statistics are not necessarily appropriate for city-level analysis. There is much debate about the quality of Google Analytics' location data, but all analysts essentially agree that there is a large degree of geographic imprecision in these statistics. Visitors are not geocoded to a street address as in the case of Google Maps. Instead, Google Analytics locations are based on visitors' IP addresses, which have a spatial precision of tens of kilometers. Commercial products used to support such geocoding, like MaxMind's GeoIP GeoLite database (http://www.maxmind.com/app/geolite), advertise an accuracy of 78 percent at the city level in the United States, within a 40 kilometer radius. A circle with this radius is over 5,000 square kilometers in area, which is much larger than the average Wisconsin county. In addition, the actual IP addresses being geocoded depend on the geographic locations of Internet Service Provider hubs where capacity is available at any given moment, rather than an individual home or office computer (Adams 2011).

In short, the imprecise nature of location information in Google Analytics can cause the visitor counts for individual cities to include large areas beyond the city boundaries, including portions of neighboring counties. To account for this geographic imprecision, visitor counts for all cities, villages, towns, and unincorporated places were tabulated to arrive at a total visitor count for each of Wisconsin's 72 counties. Iron and Florence, two sparsely-populated counties, reported no visitors, bringing the final count to 70 observations. Note that these two counties most likely have WHAIFinder visitors; it is the imprecise nature of location information in Google Analytics that gives rise to this anomaly.

Population data for Wisconsin counties was obtained from the Wisconsin Legislative Technology Services Bureau, which processes U.S. Census data for use by the Legislature and its service agencies (http://legis.wisconsin.gov/ltsb).

Broadband data from the National Broadband Map was also obtained; data collection methodologies for the National Broadband Map are complex, and a variety of broadband statistics are available. Broadband speed refers to the amount of data a user can download or upload per second, and it is usually measured in kilobits per second or megabits per second. Broadband speed determines how quickly a Web site loads or how much time it takes to download a file. For WHAIFinder, the impact of broadband speed is the time it takes to download a scanned photograph. At the county level, data on download speeds of 3 Mbps or faster were obtained. This speed is in line with the FCC's National Broadband Plan as a desired base level of access. Note that this variable measures potential access to broadband and not actual broadband use.

The primary statistical method used in this study was regression analysis. Regression analysis is similar to correlation analysis in that it quantifies the degree of similarity between two variables. However, regression analysis makes a distinction between these two variables, calling one the dependent variable and the other the explanatory variable (or independent variable). In regression, the values of the explanatory variable are used to explain the values of the dependent variable. Regression analysis provides details about the relationship between these two variables and provides a best-fit mathematical model allowing the dependent variable to be predicted by plugging in values of the explanatory variable.

A regression model involving a single explanatory variable is called bivariate regression. When more than one explanatory variable is used, it is called multiple regression. Multiple regression is used when the phenomenon being modeled is complex and cannot be accounted for using one variable alone. The ability of the explanatory variable to predict the dependent variable is typically quantified by a measure such as the R^2 ("R-square"). R^2 values range from 0.0 to 1.0, with higher values indicating greater explanatory power. Technically, the R^2 value is the proportion of the variance in the dependent variable that is explained by the explanatory variable through the regression model.

Usually a regression model is an equation for a straight line. When a linear assumption is adopted and the regression model has this linear form, the technique is referred to as linear regression. Linear regression is analogous to a trend line through a scatterplot of the dependent and explanatory variables. The two coefficients for this regression line (i.e., the slope and intercept) can be estimated using several mathematical techniques. However, in some cases there is reason to believe that the relationship is nonlinear in form. A well-documented example of this in geography is the relationship between the land area and population of cities; the area of a city increases with population but at a decreasing rate, since larger cities tend to have higher population densities (Veregin and Tobler 1997).

If a nonlinear relationship is present, a linear regression model will generally yield low explanatory power. There are various ways to overcome this problem; one of the simplest involves a transformation of the dependent and explanatory variables by taking the logarithm of their values. Since both variables are transformed, this is sometimes called a double-log regression. The log-transformed variables are then run through conventional linear regression analysis. In effect, the log transformation removes nonlinearity in the relationship to conform to the assumptions of linear regression analysis.

In this study, bivariate regression analysis was performed using the number of visitors as the dependent variable and a population value as the explanatory variable. Two separate regression analyses were conducted for this study, one using county population figures for 2010 as the explanatory variable, and the other using county population figures weighted by the percentage of the population with access to broadband with speeds of at least 3 Mbps. For ease of discussion, the two explanatory variables will be referred to as "population" and "access," respectively. Both linear and double-log regressions were performed to test the possibility of either linear or nonlinear relationships. One possible explanation for nonlinearity is higher demand in rural areas, where alternate forms of access to historic photographs and maps are more limited. This potential explanation could be tested if visitor data could be aggregated at a finer geographic scale or if counts of photograph downloads could be geocoded (located geographically) and analyzed in relation to the urban-rural footprint.

Multiple regression using population and access together in the same model was not conducted. While more complex models generally produce more explanatory power, in this case the access variable was considered as an improved version of the population variable, rather than an independent variable in its own right. The access variable was expected to deliver higher explanatory power than population because the access variable accounts for broadband access rates.

A test for "spatial autocorrelation" using Moran's *I* statistic was also conducted. Spatial autocorrelation can be thought of as a mathematical measure of geographic clustering. Positive spatial autocorrelation would indicate that neighboring observations shared similar values, while negative autocorrelation would indicate they had dissimilar values.

Regression residuals were examined to identify geographic patterns in the data. The residual value is the difference between the actual value for a given observation and the value predicted by the regression model. The sign of the residual—whether positive or negative—indicates whether the model overestimates or underestimates the value for that observation. A large residual would mean the model did not predict that observation very well. Spatial autocorrelation in residuals was tested using Moran's *I* statistic. The presence of autocorrelation in residuals can complicate the interpretation of statistical results because it violates the assumption of independence between observations.

RESULTS

Basic Statistics and Spatial Patterns of Use

Over the one-year time frame of the study (June 15, 2011 to June 14, 2012), WHAIFinder generated 21,109 pageviews; 16,497 visits; and 9,725 visitors worldwide. Figure 2 shows visitor counts summed for each Wisconsin county. The total visitor count for Wisconsin during this time period was 7,640. As anticipated, the map shows the highest visitor counts in many of the most populous counties, including Dane and Milwaukee. The Moran's *I* statistic for the distribution was not significant at the 95 percent level based on standard methods of inference testing. Spatial correlation in this context would mean that counties sharing a border tend to have similar numbers of visitors. Since significant autocorrelation is not indicated, there is no evidence of such a pattern.

Regression Analysis

For population, the double-log regression produced the best result, suggesting the data have a curvilinear relationship (Table 2). The model was statistically significant and accounted for 56.3 percent of the variance in the dependent variable. The slope coefficient was significantly different from zero and was positive, indicating the number of visitors increases with population. Since the regression model was based on the natural logs of visitors and population, it can be expressed as

$$\text{Visitors} = 0.00193 \times \text{Population}^{0.937} \quad R^2 = 0.563 \tag{1}$$

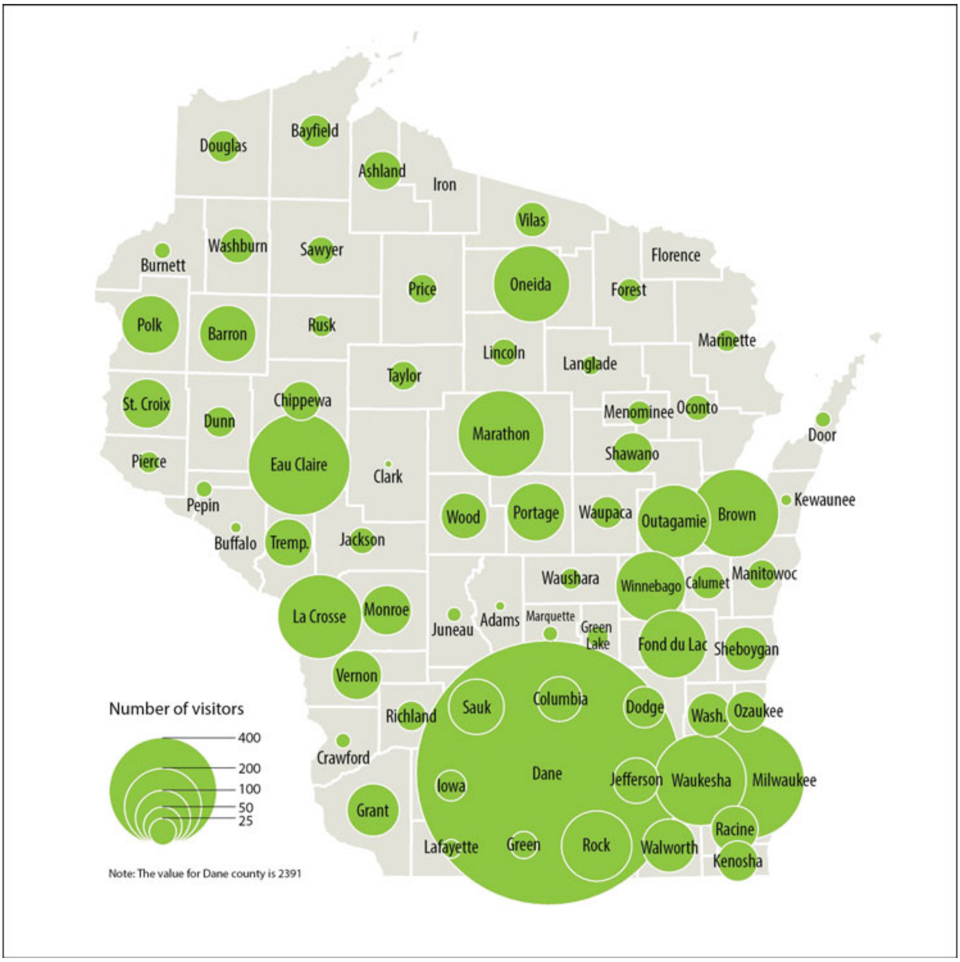

FIGURE 2 WHAIFinder visitors by county. (Color figure available online.)

The exponent 0.937 is less than one, which implies that the number of visitors increases with population and that the rate of increase declines as population gets larger.

For the access variable, the double-log regression also produced the best results (Table 3). The model was statistically significant and accounts for 58.9

TABLE 2 Regression Results, Log of WHAIFinder Visitors vs. Log of County Population

Coefficient	Estimate	t-Value	$p > \|t\|$
Intercept	−6.249	−5.807	< 0.001
Slope	0.937	9.361	< 0.001

$R^2 = 0.563$; $n = 70$; $F = 87.636$; $p < .001$.

TABLE 3 Regression Results, Log of WHAIFinder Visitors vs. Log of Population with Access to >3 Mbps Broadband

Coefficient	Estimate	t-Value	$p > \|t\|$
Intercept	−6.164	−6.092	< 0.001
Slope	0.934	9.874	< 0.001

$R^2 = 0.589$; $n = 70$; $F = 97.500$; $p < .001$.

percent of the variance in the dependent variable. The slope coefficient was positive and significantly different from zero. In this case, the regression model can be expressed as

$$\text{Visitors} = 0.00210 \times \text{Access}^{0.934}\ R^2 = 0.589 \qquad (2)$$

The equation shows that the number of visitors increases with access, but at a declining rate. Relative to model 1, model 2 predicts higher usage rates for a given value of the explanatory variable. This is as expected, since it indicates that usage rates will be higher if the pool of potential users has greater access to high-speed broadband.

The R^2 values for these two models express explanatory power in terms of the values of the log-transformed variables. While this is legitimate, the R^2 values for the original, untransformed values can also be computed from the above equations, which is more useful from a practical standpoint. The recomputed R^2 values for models 1 and 2 are 0.277 and 0.286, respectively, indicating a lower degree of explanatory power for the untransformed data. These models explain only 27.7 percent and 28.6 percent of the variance in the dependent variable.

Residuals

In the study's regression models, the residual value for a given county is the difference between the actual number of visitors for that county and the number of visitors predicted by the regression model. The authors examined residuals to identify possible geographic patterns in the data. Figure 3 is a map of residuals for model 2 (the map for model 1 is very similar). The map clearly shows that Dane county has the largest positive residual (more visitors than predicted). Other counties with large positive residuals include Eau Claire and La Crosse, and several counties in the northern parts of the state. In contrast, many of the counties along the eastern edge of the state have fewer visitors than predicted.

The authors tested for spatial autocorrelation in residuals using Moran's I statistic and found values for the two models that were not significant at the 95 percent level, meaning that significant autocorrelation was not indicated. Testing for autocorrelation in residuals is important, because the

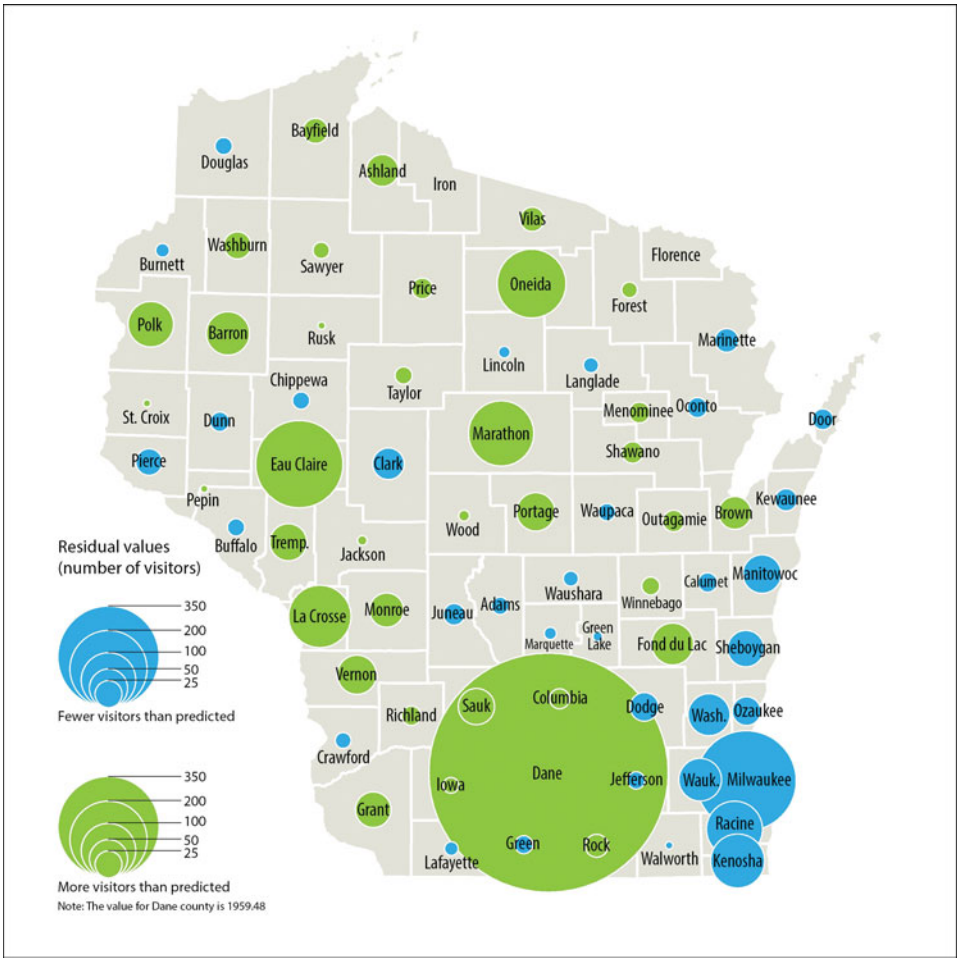

FIGURE 3 Residuals from regression model. (Color figure available online.)

presence of autocorrelation can complicate the interpretation of statistical results.

To assess the influence of Dane County on the regression results, Cook's distance (D_i) was used. This metric is based on predicted values for the dependent variable derived from the full regression model and from a partial model in which a particular data point has been removed (Cook 1977). For bivariate regression, a suggested guideline is that a D_i value greater than 0.7 indicates an "outlier": an extreme residual that can be considered for removal from the regression model (McDonald 2002). Dane County's D_i value is 5.63, several times larger than the 0.7 threshold. Removing Dane County from the model increased explanatory power significantly.

DISCUSSION

Best-fit models derived using regression methods indicate that WHAIFinder usage is positively correlated with population at the county level in Wisconsin. Weighting county population by access to broadband marginally improves the model's explanatory power, indicating that broadband access plays some role in WHAIFinder usage. However, the relatively small increase in explanatory power suggests that the particular broadband speed selected in the study (3 Mbps) might not be that meaningful in terms of the effectiveness of WHAIFinder interaction. Alternatively, this result might indicate that demand is inelastic in relation to broadband speed, due to the uniqueness of the WHAIFinder services.

Despite the generally uniform pattern, there are some isolated pockets within the state where WHAIFinder usage is higher than expected, including Dane County. High usage rates in Dane County are probably explained by the fact that this county contains numerous state agency offices and the UW-Madison campus, both of which probably contain large numbers of users. Eau Claire and La Crosse counties also have large UW campuses and similarly exhibit high usage rates. In other cases, high usage might be associated with higher levels of referred traffic due to the presence of a link to WHAIFinder on a county or city Web site. Statistics show, for example, that Eau Claire County's Web site was responsible for referring 422 visitors to WHAIFinder—over 6 percent of all referrals. There are also several relatively high-use counties in northern Wisconsin. High demand in these areas may reflect the benefits of using the online application versus driving to Madison to use the physical collection, and may also indicate a general lack of access to alternate forms of digital historic data in these areas. Increased access to the digital photograph collection by residents of more remote areas was one of the original rationales for the WHAIFinder project.

WHAIFinder usage rates were lower than expected for most counties along the eastern edge of the state. The cause of this pattern is unknown, although it has been suggested that southeastern Wisconsin is somewhat independent from the rest of the state in terms of geospatial activity (Day and Ghose 2012). In addition, several local governments in this region have made mosaics of historic imagery available on their Web sites, which might be associated with decreased demand for WHAIFinder services.

These geographic patterns have implications for the SCO's outreach efforts. The objective of the WHAIFinder project was to provide online data access for all Wisconsin residents and in all regions of the state. Results suggest additional outreach efforts could improve usage rates in underserved counties, in part by building on the successes in high-use counties. An important finding in terms of this outreach strategy is the nonlinear relationship observed between population and visitors, whereby the number of visitors increases with population but at a declining rate. At this time, the exact cause of this effect is unknown, but it may relate to higher levels of demand in rural areas. This hypothesis could be tested if visitor data could be aggregated at a finer geographic scale than counties.

This study demonstrates the utility of Web analytics methods for assessing usage rates for WHAIFinder and similar online data delivery tools. These results parallel earlier work by Jones et al., Fang, Usher, and others. The study has provided metrics of WHAIFinder usage that can serve as a baseline for future planning efforts, provide justification for resource expenditures and funding requests, and factor into ROI calculations. The study also demonstrates the value of mapping and analyzing geographic patterns, which has allowed us to identify areas where performance is lower than expected and to selectively target these areas for additional outreach and education efforts. This geographic focus, which builds on earlier examples by Tian et al. and Rodgers, provides information that can help the SCO improve and enhance its role within the state.

There are some limitations to this study. First, it has barely scratched the surface of Web analytics. Counts of visitors are useful for revealing patterns of overall access and use, they but do not quantify how many of these visitors are actually finding the photographs they need or are successfully downloading them. In the future, the authors would like to tabulate photograph downloads and segment these counts geographically to make maps of the popularity of different parts of the state. This would provide useful information for future enhancements to the system, such as identifying regions where higher-resolution scans or additional photographs from other years might be in demand. The authors are also interested in investigating user interaction with the site to determine if there are design factors that are negatively impacting usability.

A significant amount of variation remains unexplained by the regression models. There are no doubt other factors at work that affect the bivariate relationships examined between visitor counts and population. Initially, the authors had hoped that some of the additional variation might be explained by the volume of referred traffic. In practice, however, it is difficult to determine the location of origin for many referred visits, since this information is not always apparent from the Web site URL and since visitors need not reside in the county in which the referring Web site is hosted. Still, researchers such as Plaza (2009) have successfully used traffic source as an explanatory variable to account for usage rates and levels of user interaction, and we hope to include some of these approaches in future research.

Additional sources of unexplained variation no doubt arise from dimensions of the digital divide that are not considered in the study. Many socioeconomic and educational factors are known to affect the access and use of Internet technology. It might be useful to combine metrics from U.S. Census data into the regression models to account for such exogenous factors; if regression model results improved significantly, we could be more confident that we understood the reasons for low WHAIFinder usage in certain counties.

CONCLUSION

Web analytics offer a powerful approach for analyzing the effectiveness of online information delivery mechanisms. In this study, usage statistics helped quantify the level of service that the WHAIFinder application provides to the community. When coupled with census data, usage statistics allowed us to quantify and map usage patterns across Wisconsin and relate these patterns to expectations and program goals. Statistical results also facilitated identification of underserved areas where additional outreach and educational efforts are needed.

Similar methods could be used in other contexts. Web analytics make it relatively easy for library administrators to assess the value and effectiveness of Web-based information delivery tools. Even though statistics like visitor counts are basic metrics of Web usage, they still offer a great deal of useful information if analyzed in reference to specific program goals and objectives. In this way, library administrators could begin to identify factors that affect demand for online services and monitor how demand is affected by specific actions such as rollouts of new Web site features. This information could help focus resources more effectively and efficiently on those types of activities that produce the greatest impact and lead to the most engaged use.

This WHAIFinder study provides necessary baseline data for more detailed analysis, including tracking performance on specific goals and conducting ROI analysis. Generally speaking, goals can be any particular strategy or priority. For example, one might want to increase usage by a certain percentage, increase market penetration rates for certain segments of the audience, add a certain number of new users each month, or convert a target percentage of new users to returning users. Once articulated, such goals can be assessed using appropriate KPIs and analyzed using statistical methods like regression analysis. In contexts where cost recovery is a concern, such analyses can also be used to determine appropriate data access fees. In the WHAIFinder case, for example, costs might include photograph preparation and scanning, data storage, interface development and maintenance, and IT support. These figures could be used to compute a per-unit photograph download cost, by dividing total annual costs by the number of downloads on a yearly basis.

One of the more unique aspects of this study is its geographic approach, including statistical mapping and analysis of geographic patterns of usage. These methods made it possible to assess compliance with statewide program goals of data access and identify areas that appear to be underserved and in need of attention. Similar methods can be employed by library administrators to assess the geographic reach of a library's Web site and help fine-tune the delivery of information. Possible questions include the following: Where are users coming from, and does the pattern match the expectation and mission of the library? Given the locations of users, is the mix of content appropriate, or are there other types of content that should be added? Are there regions that have lower use than expected, and if so what might account for this pattern? What lessons can be learned from regions with the highest rates of use? Patterns of usage could also be coupled with socioeconomic data to gain a profile of user characteristics, which could be used to gain a better understanding of users' interests to help prioritize Web site update efforts, drive content strategy decisions, and increase user satisfaction rates.

Much more remains to be done with Web analytics, both for the WHAIFinder project and more generally in terms of online data delivery in libraries. Given the ready availability of tools like Google Analytics, studies and analyses such as this one could easily become part of the standard repertoire of all library Web site administrators. Basic analyses are easy to perform, and more sophisticated approaches will evolve naturally in response to each library's particular needs. Hopefully, this study will stimulate further thought and additional research in this area.

ACKNOWLEDGEMENTS

An earlier draft of this paper appeared in the AutoCarto 2012 Proceedings (Veregin and Wortley 2012). We gratefully acknowledge the contributions of Peter Gorman (UWDCC), Jaime Stoltenberg (Robinson Library), and other members of the project team, for development and ongoing maintenance of the WHAIFinder application. Mark Harrower, formerly at UW-Madison, was the original PI on the WHAIFinder project.

FUNDING

WHAIFinder was developed under a grant from the Ira and Ineva Reilly Baldwin Wisconsin Idea Endowment (http://www.provost.wisc.edu/baldwin.htm) awarded to the UW-Madison Department of Geography, the SCO, the Robinson Map Library (http://www.geography.wisc.edu/maplib), and UW Digital Collections Center (http://uwdcc.library.wisc.edu).

ABOUT THE AUTHORS

Dr. Howard Veregin was appointed State Cartographer in 2009 following previous positions as Director of Geographic Information Services at Rand

McNally and Associate Professor of Geography at the University of Minnesota-Twin Cities. His interests include GIS-based cartography and analysis methods, new techniques for geospatial data development and processing, and emerging applications of geospatial technology. He holds a PhD in Geography from the University of California Santa Barbara.

A. J. Wortley is Senior Outreach Specialist with the State Cartographer's Office, where he has worked since 1998. He provides technical leadership for online geospatial data catalogs and Web services, and leads the software development teams for the office's online applications including WHAIFinder. He holds a BS in Civil Engineering from UW-Madison and is also past-President of theWisconsin Land Information Association.

REFERENCES

Aabø, Svanhild. 2009. "Libraries and Return on Investment (ROI): A Meta-Analysis." *New Library World* 110:311–24. doi:10.1108/03074800910975142.

Adams, Barry. 2011. "Geotargeting Based on IP Address Is Broken." *State of Search*. Retrieved January 24, 2013. http://www.stateofsearch.com/geotargeting-based-on-ip-address-is-broken/.

Bodenhamer, David J., John Corrigan, and Trevor M. Harris, eds. 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press.

Cook, R. Dennis. 1977. "Detection of Influential Observations in Linear Regression." *Technometrics* 19:15–8. http://www.jstor.org/stable/1268249.

Day, Patrice, and Rina Ghose. 2012. "E-planning through the Wisconsin Land Information Program: The Contexts of Power, Politics and Scale." *International Journal of E-Planning Research* 1:75–89. doi:10.4018/ijepr.2012010107.

Durante, Kim, and Zheng Wang. 2012. "Creating an Actionable Assessment Framework for Discovery Services in Academic Libraries." *College and Undergraduate Libraries* 19:215–28. doi:10.1080/10691316.2012.693358.

Fang, Wei. 2007. "Using Google Analytics for Improving Library Web Site Content and Design: A Case Study." *Library Philosophy and Practice, LPP Special Issue on Libraries and Google* 1–17.

Finch, Jannette L. 2010. "Measuring Effectiveness in a Virtual Library." *Journal of Web Librarianship* 4:63–78. doi:10.1080/19322900903517466.

Goodchild, Michael F. 2004. "The Alexandria Digital Library: Review, Assessment, and Prospects." *D-Lib Magazine* 10. http://www.dlib.org/dlib/may04/goodchild/05goodchild.html.

Goodchild, Michael F., and Donald G. Janelle. 2010. "Toward Critical Spatial Thinking in the Social Sciences and Humanities." *GeoJournal* 75:3–12. doi:10.1007/s10708-010-9340-3.

Hargittai, Eszter. 2002. "Second-Level Digital Divide: Differences in People's Online Skills." *First Monday* 7. Retrieved June 11, 2013. http://firstmonday.org/issues/issue7_4/hargittai/index.html.

Jones, Casey, Sarah Giersch, Tamara Sumner, Michael Wright, Anita Coleman, and Laura Bartolo. 2004. "Developing a Web Analytics Strategy for the National Science Digital Library." *D-Lib Magazine* 10. http://www.dlib.org/dlib/october04/coleman/10coleman.html.

Kyem, Peter A. K., and James C. Saku. 2009. "Web-Based GIS and the Future of Participatory GIS Applications within Local and Indigenous Communities." *The Electronic Journal on Information Systems in Developing Countries* 38:1–16. https://www.ejisdc.org/ojs2../index.php/ejisdc/article/viewFile/584/292.

LaPointe, Evan. 2010. "A Better Definition of Web Analytics." *Atlanta Analytics*, April 22. Retrieved January 2012. http://www.atlantaanalytics.com/practicing-web-analytics/a-better-definition-of-web-analytics/.

Marek, Kate. 2011. "Using Web Analytics in the Library." *Library Technical Reports* 47.

McDonald, Barry. 2002. "A Teaching Note on Cook's Distance—A Guideline." *Research Letters in the Information and Mathematical Sciences* 3:127–8. http://www.massey.ac.nz/massey/fms/Colleges/College%20of%20Sciences/IIMS/RLIMS/Volume03/A_Teaching_Note_on_Cooks_Distance-A_Guideline.pdf.

Missingham, Roxanne. 2005. "Libraries and Economic Value: A Review of Recent Studies." *Performance Measurement and Metrics* 6:142–58. doi:10.1108/14678040510636711.

National Research Council. 1999. *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press.

Neville, Anne. 2013a. "New Broadband Map Data Shows Progress, But Work Remains." *National Broadband Map Blog* Retrieved September 10, 2013. http://www.ntia.doc.gov/blog/2013/new-broadband-map-data-shows-progress-work-remains.

———. 2013b. "NTIA Explores Broadband Availability in New Report Series." *National Broadband Map Blog* Retrieved June 11, 2013. http://www.broadbandmap.gov/blog/3009/ntia-explores-broadband-availability-in-new-report-series/.

Plaza, Beatriz. 2009. "Monitoring Web Traffic Source Effectiveness with Google Analytics: An Experiment with Time Series." *Aslib Proceedings* 61:474–82. doi:10.1108/00012530910989625.

Rodgers, Emily Puckett. 2011. "Measuring our Impact: The Open.Michigan Initiative." *Proceedings of OpenCourseWare Consortium Global 2011: Celebrating 10 Years of OpenCourseWare*. Retrieved January 24, 2013. http://deepblue.lib.umich.edu/bitstream/handle/2027.42/91308/Measuring_our_Impact_OpenMichigan_Rodgers.pdf.

State of Wisconsin. 2012. "University of Wisconsin System." Retrieved July 5, 2012. https://docs.legis.wisconsin.gov/statutes/statutes/36.

Tian, Hao, Dana J. Brimmer, Jin-Mann S. Lin, Abbigail J. Tumpey, and William C. Reeves. 2009. "Web Usage Data as a Means of Evaluating Public Health Messaging and Outreach." *Journal of Medical Internet Research* 11:e52. doi:10.2196/jmir.1278.

Turner, Steven J. 2010. "Web Site Statistics 2.0: Using Google Analytics to Measure Library Web Site Effectiveness." *Technical Services Quarterly* 27:261–78. doi:10.1080/07317131003765910.

Usher, Oli, Oana Sandu, and Lars Lindberg Christensen. 2012. "How Do We Know What Works?" *CAPjournal* 12:9–12. http://capjournal.org/issues/12/lowres.pdf#page=9.

Veregin, Howard, Peter C. Gorman, Jaime Stoltenberg, A. J. Wortley, and Michael Bricknell. 2010. "An Open Source Web Application for Historic Air Photo Display and Distribution in Wisconsin." *Proceedings, AutoCarto 2010*. http://www.isprs.org/proceedings/XXXVIII/part4/files/Veregin.pdf.

Veregin, Howard, and Timothy Kennedy. 2012. "Online Information Dissemination at the Wisconsin State Cartographer's Office Using Map Services and APIs." In *Online Maps with APIs and Map Services*, edited by Michael Peterson, 279–98. Berlin Heidelberg: Springer.

Veregin, Howard, and Waldo R. Tobler. 1997. "Allometric Relationships in the Structure of Street-Level Databases." *Computers, Environment and Urban Systems* 21:277–90. http://dx.doi.org/10.1016/S0198-9715(97)00012-4.

Veregin, Howard, and A. J. Wortley. 2012. "Analysis of Web-Usage Statistics to Evaluate the Effectiveness of Online Maps for Community Outreach." *Proceedings, AutoCarto 2012*. http://www.cartogis.org/docs/proceedings/2012/Veregin_Wortley_AutoCarto2012.pdf.

Wang, Yong, Phillip Lai, and Daniel Sui. 2003. "Mapping the Internet Using GIS: The Death of Distance Hypothesis Revisited." *Journal of Geographical Systems* 5:381–405. doi:10.1007/s10109-003-0117-9.

Web Analytics Association. 2008. "Web Analytics Definitions." Retrieved January 23, 2013. http://c.ymcdn.com/sites/www.digitalanalyticsassociation.org/resource/resmgr/PDF_standards/WebAnalyticsDefinitions.pdf.

Wood, Fred B., Elliot R. Siegel, Eve-Marie LaCroix, Becky J. Lyon, Dennis A. Benson, Victor Cid, and Susan Fariss. 2003. "A Practical Approach to E-Government Web Evaluation." *IT Pro* 5:22–8. doi:10.1109/MITP.2003.1202231.