All that glitters isn't gold: The complexities of usage statistics as an assessment tool for digital libraries

Abstract

Purpose

Digital collection assessment has focused mainly on evaluating systems, metadata, and usability. While usage evaluation is discussed in the literature, there are no standard criteria and methods for how to perform assessment on usage effectively. This article asserts that usage statistics have complexities that prohibit meaningful interpretation and assessment. The authors aim to discover the problems inherent in the assessment of digital collection usage statistics and propose solutions to address such issues.

Approach

This article identifies and demonstrates five inherent problems with usage statistics that need to be addressed when doing assessment for digital collections using the statistics of assessment tools on local digital repositories. The authors then propose solutions to resolve the problems that present themselves upon such analysis.

Findings

The authors identified five problems with digital collection usage statistics. Problem one is the difficulty of distinguishing different kinds of Internet traffic. Problem two is the lack of direct correlation of a digital item to its multiple URLs, so statistics from external web analytics tools are not ideal. Problem three is the analytics tools' inherent bias in statistics that are counted only in the positive way. Problem four is the different interaction between digital collections with search engine indexing. Problem five is the evaluator's bias towards simple growing statistics over time for surmising a positive usage assessment. Because of these problems, statistics on digital collections do not properly measure a digital library's value.

Value

This article identifies five problems that need to be addressed before a meaningful assessment of digital collection usage statistics can take place. The article ends with a call for evaluators to try to solve or mitigate the stated problems for their digital collections in their own evaluations.

Keywords: Digital Collection Assessment, Digital Collection Evaluation, Usage Statistics, Web Analytics, DSpace, Institutional Repository, Digital Repository

Introduction

Beginning in the 1990s, institutional digital libraries and collection arose across a variety of specialties and geographic locations. (For the purposes of this article, *digital libraries* and *digital collections* refer to the digitized and born-digital items found in institutional repositories and special collections.) This variety led to a proliferation of methods, both theoretical and practical, for digital collection assessment. These methods are focused on how libraries created these collections, such as the systems used, metadata implemented, and usability of the website. What is lacking from the current literature is a standardized model for usage assessment of these digital collections, or evaluating how much the collection is actually being utilized by online patrons. Without this model, digital collection curators are not able to evaluate if their digital collections are increasing, maintaining, or losing their value for the institution. The digital collection management field has not yet been able to address a standardized method to evaluate usage statistics.

The authors attempted to evaluate two digital collections' usage over time and found five problems with usage statistics when looking at views or page views. This article discusses and reviews the inherent problems with views and page views statistics as an assessment tool in detail. The authors then showed why looking at user sessions is a more effective assessment for usage statistics specifically because sessions avoid many of the problems experienced with views or page views. The authors also argue that the internal statistics engine of DSpace is insufficient for effective usage statistics evaluation, and that digital collection evaluators should use Google Analytics to capture usage statics. The authors also provide suggested metrics in Google Analytics to evaluate digital collections.

Literature Review

With the proliferation of digitization projects, there is an increasing need for digital collection assessment (Xie, 2006). It is difficult, however, to define what a good, quality digital library is because assessment standards depend on many aspects (Heradio *et al.*, 2012; Moreira *et al*., 2009). The difficulty in defining what a quality digital library is has led to the development of many facets of digital library assessment, including database structure, network architecture, protocol interoperability, usability, information architecture, etc. (Tsakona *et al.*, 2004; Tsakonas & Papatheodorou, 2011; Heradio *et al*., 2012).

Since there are no standard techniques and criteria for digital library or digital collection evaluation, previous research applied evaluation criteria from traditional libraries, human-computer interaction, digital repository system performance, and digital technologies (Xie, 2008). Most relevant studies adopt traditional approaches and criteria to examine common features of digital collection system, concentrating on evaluating the impacts of digital collection and systems on the scholar populations (Saracevic, 2000; Fuhr *et al.*, 2001; Tsakonas & Papatheodorou, 2008; Zhang, 2010).

Traditionally, an effective way to assess a diverse physical library collection was to focus on two indicators, capacity and usage (Borin & Yi, 2008; Borin & Yi, 2011; Henry *et al*., 2008). The transition period from primarily print formats to primarily digital formats requires researchers to take account of this shifting landscape into collection evaluation (Borin & Yi, 2008). An influential publication on selecting collections for digitization written by a Harvard task force (Hazen *et al*., 1998) included the usage assessments of analog and digitized collection as a criterion, considering how much the collection is used and what the nature is of the use. Marchionini (2000) also suggested applying existing techniques and metrics to the evaluation of digital libraries, including circulation, collection size, and patron visits. Almost ten years later, in 2009, Ooghe and Moreels found that usage is still a criterion for digitized collections. Arendt and Wagner's research in 2010 was a practical example of measuring digital usage, as it utilized Google Analytics to analyze geographical location of users of a digital collection.

In 2012, Schlosser and Stamper stated in their study that, although a number of articles have been written regarding usage measuring of library websites in general, almost two decades into library collection digitization programs there is not yet a significant body of literature regarding usage measurement of digitized collections. Schlosser and Stamper argue that librarians working with born-digital and digitized collections had fallen behind in measuring the use of their collections. However, increasing the usage of digital collections has always been a major goal for digital collection managers. Since usage measuring and evaluation of digital resources are challenging, Schlosser and Stamper requested that the libraries that do collect usage statistics on digital collections should share the methods and results with peers to help foster an environment where such data are collected and used.

Reviewing the literature, the evaluation of digital libraries has both theoretical and practical aspects (Marchionini, 2000). There is a gap, however, between evaluation theorists and evaluation practitioners since the developed models and concepts of digital library evaluation are not always applied in practical studies (Saracevic, 2004). Practical evaluation activities of digital libraries need to answer three essential questions: why evaluate, what to evaluate, and how to evaluate (Saracevic, 2004; Fuhr *et al*., 2007). Tsakonas and Papatheodorou (2008) expressed a similar concern, emphasizing that an evaluation activity depends on several parameters, including the purpose of the research, the macroscopic view of the evaluation, and the object of the evaluation.

Other usage assessment tools in the library world such as COUNTER have been created to deal with the problem of gathering usage statistics from third party applications (Blecic, Fiscella, & Wiberley, 2007; Bernon, 2008; Stewart, 2011). Such tools are designed specifically for subscription-based electronic resources instead of institutionally unique digital resources. That said, applications such as COUNTER will not work with digital collections because of the differences in third party applications and digital collections. One difference between digital collections and subscription based electronic resources is the issue that staff adding to the collection can polute the usage statics in a digital collection, while in a subscription based electronic resources this is not a big problem because staff are not touching every item that is entered into each database. This staff use causes a number of problems that will be discussed in more detail below, but they are problems that are not faced by applications such as COUNTER. Because digital collections are typically open access collections, there is a greater chance of accidental use and a greater need to distinguish between different kinds of usage. With electronic resources, because of the nature of third party applications, their resources cannot be used by people who are not part of the organization that has

paid for the resources, greatly reducing the instances of accidental use. The often open access digital collections also rely heavily on how easy it is for search engines to crawl their content in order to make the materials discoverable. With third party systems, they usually have a self-contained search mechanism, or the materials are discoverable through a federated search. With these stated differences, digital collections act more like websites than electronic resources from third party systems, and web analytics software is best suited to evaluating the collections.

In a series of articles about web analytics tools, Marek (2011 Chapter 1) insisted that librarians must adopt a return-on-investment mindset to evaluate library services, including digital collections, to demonstrate digital resources made available through the website are constantly responding to trends and user needs. In a recent study of digital library assessment, Kelly (2014) concluded that studying web analytics has become one of the most prevalent techniques in digital library evaluation. There has been an exponential growth of web analytics tools in the past decade (Marek, 2011 Chapter 1). Many libraries have been using Google Analytics since it was released to the public in mid-2006 to measure their library website performance and to analyze user behaviors. This popularity can be traced to Google Analytics' free availability, tremendous features, and ease of use (Fang, 2007; Marek, 2011, Chapter 1). Unlike other web analytics tools using server log files, Google Analytics collects information by inserting simple JavaScript codes into pages, which enables the tool to capture detailed demographic information that log files do not normally provide (Turner, 2010). However, the raw data and factual statistics from analytics tools only tell a portion of the story (Kaushik, 2010). Librarians and evaluators have to find their own ways to comprehend such information and to interpret the data by combining relevant analytics metrics and measurable indicators.

Methodology

Missing from the literature is a common standard to evaluate usage of digital collections. This lack of a common standard is due to some complex issues with web or usage statistics being applied to a digital library. While trying to develop an evaluation model for the local digital collections, the authors encountered five inherent problems that prohibit the assessment. By pointing out these problems that are commonly encountered in the process of digital collection evaluation, the authors also proposed solutions to address these issues.

Two different copies of DSpace housing different types of digital collections were used for this study. DSpace is a content management system created for institutional repositories that is being used by over 1,000 organizations worldwide (DSpace, 2015). The authors tried different methods of assessing usage statistics in order to identify the health of the collections and what audiences the collections served.

The first DSpace instance is the authors' institutional repository, which housed 35,751 titles in 2015. Of those items, 20,082 were electronic theses and dissertations (ETDs). The rest of the collection is made up of faculty research, scholarly output from different institutes on campus, and a few digitized collections. This collection will be referred to as IR DSpace throughout the article. The second DSpace instance is a digital repository of digitized special collections from the same

university. In 2015, the collection contained 124,816 titles. The collection contains archival materials such as correspondence, photographs, and newspapers. This collection will be referred to as SC DSpace throughout the article. Both DSpace instances were connected to Google Analytics for three years, from 2012 to 2014, and those usage statistics were used for this evaluation.

Google Analytics Terminology

Before delving into the specifics of usage statistics, the authors would like to clarify the Google Analytics terminology that will be used throughout the article.

Users and Session(s)

Just as a physical library tracks how many patrons walk through the doors, Google Analytics uses the ***Users*** and ***Active Users*** metrics to track user interaction with a site (Google, 2015). Site users take part in "sessions" with the elements of the website. Analytics begins timing these sessions from the moment a user arrives on the site (Google, 2015). In a physical library, the equivalent of a session would be the entire time that someone was inside that library. If that person left the building, that would be the end of the session. A new session would begin when they walked back into the building. Alternatively, if someone fell asleep in the building for at least 30 minutes, that would also be the end of the session. The new session would begin when they woke back up.

Pageviews

Each instance of a page being loaded or reloaded in a browser is measured by ***Pageviews*** (Google, 2015). This would be the physical equivalent of counting every page read in a book or in a library. One person could sit down and read a 300 page book, and that would count at 300 page views. That same person could look at the first page of 300 books, and that would also count as 300 page views.

Bounce Rate

A very important metric in tracking digital collection usage is the bounce rate. "***Bounce Rate*** is the percentage of single-page sessions (i.e. sessions in which the person left your site from the entrance page without interacting with the page)." (Google, 2015). In a physical library, the equivalent would be someone walking in the door, realizing they were in the wrong building, and walking right back out. Their brief stay would still count on the gate count, as the visit still counts in usage stats, but the visit would not be substantial use of the building.

Inherent Problems in Usage Assessment when using Views or Pageviews

Digital collection usage statistics can be evaluated from two different perspectives. Either the usage can be viewed from the perspective of individual users in terms of sessions or from the

perspective of the item in terms of views or page views. The authors found the following problems with looking at usage from views or page views.

Problem 1: Distinguishing Different Kinds of Traffic

A problem with both external web statistics software and content management systems' (CMS) statistics engines is that they do not distinguish between staff use and patron use. While Google Analytics does have an option to restrict certain IP ranges, for academic institutions staff usage and patron usage may come from the same range of IPs. Staff working on digital collections can block the IP address for their computers, but depending on how they connect to the network, their machine may be issued a new IP address. One way to address this issue when analyzing the numbers is to develop a filter using the browser, the operating system, and the network that matches the staff computers so that their usage numbers can be removed from the statistics. However, this method cannot be 100% accurate. The alternative way to handle the problem is to make sure staff are aware of their own effect on the statistics and take steps to reduce that effect in reports. At the authors' institution, Google Analytics is attached to DSpace. Staff usage can be distinguished because staff work in pages that have "/admin" in the URLs, so those views can be removed manually from the total usage count because patrons cannot access the admin pages.

[Table 1]

Even though patrons cannot access the /admin pages, a lot of non-admin page use can be linked to staff use by looking at sessions and their network.

[Table 2]

In tables 1 and 2, removing obvious staff use makes it look like the SC DSpace gets nearly twice the use that the IR DSpace gets, but the situation is reversed when the data is viewed though sessions. This shows that staff use is difficult to remove in page views, but can be removed to an extent when viewing the data through sessions. Removing the staff use by network is only possible by using an external web analytics software like Google Analytics as DSpace's native interface does not capture enough information about the usage to know what network a user has come in from.

Another traffic issue is that the CMS statistics engines do not distinguish between accidental traffic and meaningful traffic without using external tracking software. Digital collection evaluators cannot simply look at the numbers of a collection and compare them alone. Evaluators have to take other statistics into consideration such as "Pages per Session", "Average Session Duration", and "Bounce Rate". These kinds of statistics can help evaluators determine the quality of usage, but such statistics are typically only available from external analytics software. Content management systems' internal statistics engines usually only count views and downloads. This means that if usage statistics are going to be used and evaluated based on quality of usage, an external web statistics software must be used. External web analytics software requires some technical experience to install, and some level of training to interpret after the point. Ideally, the digital library community can develop a few library specific usage assessment metrics and integrate the metrics

into content management systems for the convenience of digital collection curators to evaluate their collections. Until then, Google Analytics can fill in this gap.

[Table 3]

The above table shows that the bounce rate has a profound effect on the page views. These accidental bounces, when removed, reveal a more accurate number. The bounce rate is not a metric used by DSpace, meaning the only way to get it is to use an external web assessment tool. However, removing the bounce rate does not remove staff use. In 2014, the SC DSpace's collection curators started batch loading collections. This uploading process resulted in a surge in page views that were solely staff use, making the collection look healthier than it was, even with the bounced views removed.

Problem 2: Items May Have Multiple URLs

Another problem with usage statistics in digital collections is that when an external analytics software is used, it may identify a number of different URLs per digital item. For example, in DSpace, there is a URL for the main site, the individual collection pages, and within each item exists a different URL for the short metadata record, the full metadata record, and the attached item itself. All of these URLs refer to the same item. Any usage on any of these URLs should be counted as usage to this one item. An example is listed below.

Short Record URL: [insert URL]

Full Record URL: [insert URL]

Item URL: [insert URL]

Handle URI: [insert URL]

In DSpace's internal statistics software, it identifies these URLs as connected, but it also does not record the other statistics that are useful to evaluate the quality of usage. Google Analytics sees these as individual URLs and does not link them to the same item. Therefore, it is currently impossible to take DSpace usage from a system like Google Analytics and identify usage per item. Usage per item rather than usage per URL is more reasonable when interpreting statistics. DSpace does track usage per item but does not provide the extra information that Google Analytics does. Because of this, even though Google Analytics cannot provide a usage per item statistic, it is still the better tool for evaluating the quality of use on a whole digital collection or whole digital library.

Problem 3: Bias in Usage Statistics

Because view and page view statistics are taken on URLs instead of items, there is another problem inherent with usage statistics – usage statistics are only counted in the positive. Ideally, to do a meaningful evaluation the items that received usage should be compared to the number of items that receive no usage. Usage statistics do not show how many items or URLs received no usage. As a result, large collections will appear to have more usage than smaller collections simply

because they have a greater number of URLs, even if the percentage of usage per item is lower than the smaller collection. For example, in a collection of 10,000 items where each has three URLs, each URL getting one view in one year will results in 30,000 views. However, in a smaller collection of 10 items where each item has three URLs, each URL receiving 10 views within a year will results in 300 views for the whole collection. The absolute usage statistics of the smaller collection looks much worse than for the larger collection, but comparing to the size of the collection, the smaller collection is more valuable.

A few heavily used items may skew the results to be more positive than they should be, creating a bias towards those items and collections. For example, in a collection of 100 items where most items get 3 views in a year but one item gets 4,000, that single item with 4,000 views skews the total number of views to look more positive for the whole collection. The collection may look like it is receiving 4,300 views at first glance, but upon that one item's removal, the total views come down to 300. One item receiving 4,000 views over a couple of months can make it seem like the whole collection is suddenly more popular when in reality it is a fluke. A collection with a few heavily used items is not going to evaluate as well as a collection that has more generalized usage, one can assume. So, it is important to measure the breadth of usage in the collection (how many items in the collection are being used) rather than just the total number of hits the collection receives.

In the IR DSpace, the second most-viewed item has received 4,712 views (as reported by DSpace's native statistics engine) since it was put up in 2012, but the item itself is an unavailable, embargoed ETD. Google Analytics shows the item has only received 113 page views since it was put up. If the collection was being evaluated using the DSpace interface, this item would skew the results as it has gathered four thousand views despite not being available. This example shows that when looking at "views" or "page views" a few items can skew the results of a whole collection to look more positive than they are.

Problem 4: Different Kinds of Collections Act Differently with Search Engines

The length of time a collection is up can affect the usage statistics. Search engines get better over time at identifying information seekers' interests in the content, so fewer visits get to the collection by accident. Search engines also constantly update their algorithms to produce more accurate results, therefore an older collection should have different accidental usage than a newer collection. This results in some collections getting less overall use, but the use they do get is higher quality over time. In order to address this issue, the Bounce Rate can be used in evaluation. The lower the Bounce Rate, the more high quality use is reaching the collections. Bounce Rate is the percentage of single-page sessions where the users do not interact with the page at all (Google, 2015). The Bounce Rate is a measure of accidental usage because a URL is viewed but the user backed out and only spent a few seconds on the page. For digital collections, Bounce Rate is particularly useful since getting to an item usually requires the user to click on the item from a page, meaning that users have to interact with a page to get the information they want. Bounce rates from the two systems the researchers investigate are shown in Table 4.

[Table 4]

It is clear from Table 4 that the Bounce Rate for both collection is trending down, which means that they are getting higher quality use despite possibly losing total number of views. While Bounce Rate can help evaluators start to understand how effective their collections are to their patrons, it can also be difficult to understand. A low bounce rate is good until it is too low. A bounce rate below 25% can mean the tracking software may not be set up correctly (Peyton, 2014).

Collections that are complete and not being updated (called static collections) will be treated differently than actively growing collections by search engines. Search engines prioritize websites that update frequently (Google, 2010), which means growing collections will get prioritized over static collections and can receive more usage as a result. Usage for static digital collections tends to go down from year to year as the collection reaches its target audience. Someone using a digital collection once might never need to revisit the collection, so from year to year usage could drop off. This drop off can be confusing to digital collection evaluators as it looks like the collection is failing and in need of more marketing, but in fact the collection has reached its intended audience and may have fulfilled the majority of its purpose.

The time period for selection also makes a big difference in the results. The time of year that statistics are taken can make usage statistics look better or worse than they actually are. The smaller the time frame, the more likely it will make the collection either look like it is performing better or worse than it really is. However, gauging usage across the entire life of a collection is meaningless unless there is something to compare it to. For example, if a collection up for three years has 100,000 views, it is difficult for digital collection evaluators to determine if 100,000 views is successful or not without a measure. Ideally, evaluators need to break up the time period into even segments and compare the collection's usage across those segments. The problem is identifying the length of the segments.

The fluctuations of the academic calendar make even using a month as an evaluation time segment problematic. A digital collection may receive more than normal usage in the weeks before finals of an academic semester, but receive less than normal usage in the first month of the summer break. There are two ways of handling this problem: either by comparing the collection to itself in the same month the year previous or by increasing the time segment for evaluation. The ideal time frame for evaluation should be one full year, but this means that collections have to be up for a minimum of two years before they can be evaluated.

Problem 5: Human Bias towards Positive Results

Digital collection evaluators hope to have statistics to show that their collection improves over time. They want simple numbers that they can interpret to mean what they already want. Evaluators are subject to confirmation bias, meaning they will tend to read statistics as conveying what they want them to mean. This bias means that often people focus on total number of views of a collection over time. The number will always get bigger, so it always looks like the collection is improving, and it always appears impressive. However, these total numbers do not necessarily tell evaluators the number of quality visits or distinguish between bounced traffic and regular traffic. It is the number administrators and donors want to hear, but it is not a valid number for an effective evaluation.

DSpace's internal statistics engine caters to the human bias toward positive results by only providing very limited statistics at the main level and limited statistics at the community level. DSpace does provide "Total Visits per Month" for seven months' worth of statistics. After that, it provides "top country views" and "top city views". It is not possible from the user interface to select a period of time to review, and all the statistics are kept in terms of views instead of sessions, meaning that staff use cannot be removed from the results. DSpace also does not track the quality of use (by tracking things like Bounce Rate), so any usage at all will count toward this total. The collections in DSpace appear to be more heavily used than they really are. While the statistics are enough to provide large numbers to library administration, they lack the complexity to use for evaluation purposes.

Usage statistics of digital collections are complicated, and numbers that decrease will likely be discouraging, even if the decreased usage the collection is getting is more meaningful. The final problem with usage statistics is that evaluators are tempted to resist numbers or trends they perceive as negative instead of pausing to understand the context of the numbers.

Proposed Solution: Counting Sessions by Location and Network

After addressing all the problems with page views, the authors decided to look at the collections from the perspective of the users by looking at sessions. This method looked promising as many of the page view problems were no longer an issue. Staff use was easier to identify and remove by looking at the Sessions by Network. For the evaluation of sessions by network, the total number of sessions was recorded and the total number of sessions from the institution network were recorded. The institution network sessions were subtracted from the total sessions to get the sessions from other networks.

[Table 5]

[Table 6]

Using the network to remove the staff use made it clearer how much growth in usage was due to increased use by patrons and how much of it was due to increased production by the two organizations putting information into the systems. The problem in Table 1 where the SC DSpace had started generating more page views from staff use does not show up as a problem in Table 6. The number of sessions from outside the institution network in 2014 felt reasonable. This number is far more valuable to the SC DSpace administrators as it gives them the equivalent of a digital gate count and it shows consistent growth.

Interestingly, in Table 5, the IR DSpace lost non-staff usage from 2013 to 2014. The authors looked at the session data in terms of location to try to determine why that usage decreased. First, they recorded the total number of sessions and removed the home country sessions. This was the number used for the international sessions. The home state statistics were removed from the home country, the home city statistics were removed from the home state. This was done to show the audiences independently in order to more clearly show growing or shrinking audience.

The staff use in both Table 7 and Table 8 is limited to the usage in the Home City. The authors realized they could do a custom report by adding a secondary dimension to the Google

Analytics Location Report to combine the sessions by network and the sessions by location in order to remove the staff use from the home city.

[Table 7]

[Table 8]

In Table 7, the usage inside the home state (discounting the home city) clearly increased, while it was also clear that the greatest loss of audience was from the international sessions from 2013 to 2014. In Table 8, the SC DSpace has quickly gained a larger audience in the home state. Considering the collection's scope is regional in nature, this is a good sign for the health of the collection. Both collections showed a small audience gain in the home city (after removing staff use). The home state audience has apparently consistently grown with both collections.

The institution network usage by location and the institution network usage report in Tables 5 and 6 produced different numbers. This difference is because the sessions were limited by location in Tables 7 and 8 whereas Tables 5 and 6 were only limited by network. Because the authors' institution is a broad organization, the institution network exists outside the city. Therefore, the method used in table 7 and 8 is the most accurate way of removing staff use from sessions.

Conclusion

Digital collections are typically put online to be used, and as such it is vital to develop a method to evaluate usage statics effectively. Developing good methods that take into account the five problems addressed above should help digital collection evaluators identify problems with their collections or make administrative decisions about collections or how to market them. Without good digital collection usage statistics metrics, administrators are in the dark about the impact and value of their digital collections. Consequently, being ignorant of such impact and value makes budgeting and resource allocation nothing more than a guessing game.

Evaluating usage based on views or page views can be problematic because it is difficult to remove staff use and connect the URL usage to the items. Instead, institutions should look at their usage statistics through sessions or user perspective. With the method provided, audience evaluation can be made easier and allow administrators to put their resources towards collections that have a greater impact. Digital collection evaluators need to connect their resources to Google Analytics and learn how to use web analytics software.

Each digital collection evaluator should be looking for ways to address the proposed problems with usage statistics on their unique system. The authors suggest, after Google Analytics has been connected to a digital library for two years, that those who are administrators of digital collections in DSpace attempt to do the same Sessions by Network and Location evaluation suggested in this article to see if the evaluation is helpful for their institution's needs.

Limitations and Further Research

Even when looking at sessions instead of views or page views, quality users or sessions are still difficulty to quantify. For example, a collection could have 100,000 people glance at it but that does not mean as much as one researcher using the collection to write a book. In a physical library or archive, this researcher is visible by sitting in a reading room, requesting items from interlibrary loan service, or in some way interacting with librarians or archivists. Online, this quality user gets muddled with the others. Methods for identifying high quality use versus casual use need to be developed. Moreover, this evaluation was done only on DSpace. It is thus not clear if this evaluation method would work on other collection management systems.

The digital library community needs to develop its own metrics to evaluate digital collection usage that takes into account how different digital collections are from business websites, news websites, or blogs, which typically only care about the total number of views from month to month. For digital collections, it is not just about views, but about quality usage. Perhaps more important than developing common metrics is the need for the digital library community to become better at sharing experiences of how their institutions evaluate digital collections usage statistics. An informal discussion at conferences and conventions over coffee or a meal can help bridge differences and help the community solve these problems.

Reference

Arendt, J. and Wagner, C. (2010), "Beyond description: Covering web site usage statistics into concrete site improvement ideas", *Journal of Web Librarianship,* Vol. 4 No. 1, pp. 37-54.

Bernon, J. (2008). Why and how to measure the use of electronic resources. *Liber Quarterly*, Vol. *18* No. 3/4, pp. 459-463.

Blecic, D.D., Fiscella, J.B., and Wiberley, S.E. (2007). Measurement of use of electronic resources: Advances in use statistics and innovations in resource functionality. *College & Research Libraries*, Vol. 68 No. 1, pp. 26-44.

Borin, J. and Yi, H. (2008), "Indicators for collection evaluation: A new dimensional framework", *Collection Building*, Vol. 27 No. 4, pp. 136-143.

Borin J. and Yi, H. (2011), "Assessing an academic library collection through capacity and usage indicators: Testing a multi-dimensional model", *Collection Building,* Vol. 30 No. 3, pp. 120-125.

DSpace (2015), "Top reasons to use DSpace", available at http://www.dspace.org/why-use (accessed 31 December, 2015).

Fuhr, N., Tsakonas, G., Aalberg, T., Agostic, M., Hansen, P., and Kapidakis, S. (2007), "Evaluation of digital libraries", *International Journal on Digital Libraries,* Vol. 8 No. 1, pp. 21-38.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., and Sølvberg, I. (2001), "Digital libraries: A generic classification and evaluation scheme", In *Research and Advanced Technology for Digital Libraries*, Springer, Berlin Heidelberg, pp. 187-199.

Google (2010), "Search engine optimization starter guide", available at http://static.googleusercontent.com/media/www.google.com/en//webmasters/docs/search-engine-optimization-starter-guide.pdf (accessed 29 February, 2015).

Google (2015), "Bounce rate: Learn what a bounce rate is, and how to improve it", available at https://support.google.com/analytics/answer/1009409?hl=en (accessed 29 February, 2015).

Hazen, D., Horrell, J., and Merrill-Oldham, J. (1998), "Selecting research collections for digitization", available at http://www.clir.org/pubs/reports/hazen/pub74.html (accessed 29 February, 2015).

Henry, E., Longstaff, R., and Van Kampen, F. (2008), "Collection analysis outcomes in an academic library", *Collection Building*, Vol. 27 No. 3, pp. 113-117.

Heradio, R., Fernandez-Amoros, D., Cabrerizo, F. J., and Herera-Viedma, E. (2012), "A review of quality evaluation of digital libraries based on users' perceptions", *Journal of Information Science*, Vol. 38 No. 3, pp. 269-283.

Marchionini, G. (2000), "Evaluation digital libraries: A longitudinal and multifaceted view", *Library Trends*, Vol. 49, pp. 304-333.

Moreira B. L., Goncalves M. A., Laender, A. H., and Fox, E. A. (2009), "Automatic evaluation of digital libraries with 5SQual", *Journal of Informetrics*, Vol. 3 No. 2, pp. 102-123.

Ooghe, B. and Moreels, D. (2009), "Analysing selection for digitization: Current practices and common incentives", available at http://www.dlib.org/dlib/september09/ooghe/09ooghe.html (accessed 30 February, 2015).

Peyton, J. (2014), "What's the average bounce rate for a website", available at http://www.gorocketfuel.com/the-rocket-blog/whats-the-average-bounce-rate-in-google-analytics (accessed 30 February, 2015).

Saracevic, T. (2000), "Digital library evaluation: Toward evolution of concepts", *Library Trends*, Vol. 49 No. 2, pp. 350-369.

Saracevic, T. (2004), "Evaluation of digital libraries: An overview", available at http://www.scils.rutgers.edu/~tefko/DL_evaluation_Delos.pdf (accessed 30 February, 2015).

Schlosser, M. and Stamper, B. (2012), "Learning to share: Measuring use of a digitized collection on Flickr and in the IR", *Information Technology and Libraries*, Vol. 31 No. 3, pp. 85-96.

Stewart, C. (2011). Keeping track of it all: The challenge of measuring digital resource usage. *Journal of Academic Librarianship*, Vol. 37 No. 2, pp. 174-176.

Tsakonas, G., Kapidakis, S., and Papatheodorou, C. (2004), "Evaluation of user interaction in digital libraries", in Agosti, K. and Fuhr, N. (Ed.), *Notes of the DELOS WP7 Workshop on the Evaluation of Digital Libraries*, Padua, Italy.

Tsakonas, G. and Papatheodorou, C. (2008), "Exploring usefulness and usability in the evaluation of open access digital libraries", *Information Processing and Management,* Vol. 44, pp. 1234-1250.

Tsakonas G. and Papatheodorou, C. (2011), "An ontological representation of the digital library evaluation domain", *Journal of the American Society for Information Science and Technology,* Vol. 62 No. 8, pp. 1577-1593.

Xie, H. I. (2006), "Evaluation of digital libraries: Criteria and problems from users' perspectives", *Library and Information Science Research,* Vol. 28 No. 3, pp. 433-452.

Xie, H. I. (2008), "User's evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment", *Information Processing and Management,* Vol. 44, pp. 1346-1373.

Zhang, Y. (2010), "Developing a holistic model for digital library evaluation", *Journal of the American Society for Information Science and Technology,* Vol. 61 No. 1, pp. 88-110.