Assessment of E-Resource Usage at University of Massachusetts Amherst: A MINES for Libraries® Study Using Tableau for Visualization and Analysis

Rachel Lewellen, Assessment Librarian, University of Massachusetts Amherst

Terry Plum, formerly Assistant Dean, Simmons School of Library and Information Science (retired)

Introduction

As academic libraries spend an increasing amount of their budgets on electronic resources, librarians seek to find out who is using the electronic resources and why, so libraries can provide better service. One tool is COUNTER,¹ which sets standards for recording and reporting usage of networked electronic resources. From COUNTER-compliant data-usage reports from subscribing vendors, libraries receive the number of successful full-text article requests by month and journal, total search requests by platform or database, successful section requests by e-book, and various title request reports, among other reports.² As described by Emery and Stone,³ there are a number of tools for analyzing COUNTER data in addition to those in an electronic resource management system, including vendor-supplied applications like ProQuest 360 Counter,⁴ EBSCO Usage Consolidation,⁵ and consortial tools like Journal Usage Statistics Portal.⁶ With these data and tools, performance indicators such as downloads per FTE user, cost per search, cost per view, cost per FTE user, and other metrics can be generated.⁷

However, as Marcum and Schonfeld note, as useful as the COUNTER-compliant activity counts are for library usage data and performance statistics,

More granular data on individual users’ activities can afford greater opportunities to analyze needs and develop or optimize services. But considerations for user privacy can make librarians uncomfortable with such granular data, which has to some degree impeded our ability to establish the types of personalized services that are skyrocketing on the consumer internet.⁸

This paper presents a complementary methodology to COUNTER, giving the greater granularity while maintaining user anonymity and privacy. Reported here are the findings of the second, yearlong, assessment of electronic resources at the University of Massachusetts Amherst, using the Association of Research Libraries (ARL) MINES for Libraries methodology.⁹ MINES is an online, transaction-based, point-of-use, intercept web survey methodology that collects data on the patrons’ purpose of use of electronic resources and on the demographics of users. This methodology helps to measure the impact of library services and to identify opportunities to serve faculty and students more effectively, which as Marcum and Schonfeld note are the desired goals for assessment.¹⁰ The MINES data gives a picture of users and usage that does not replace COUNTER data but gives a more complete and deeper picture.
The University of Massachusetts implemented MINES twice, in 2008–2009 and 2013–2014. This paper compares two implementation methods for a point-of-use, intercept survey launched at the EZproxy server: (1) randomly chosen two-hour sessions and (2), an every-Nth-user systematic methodology. The 2008–2009 survey used 24 two-hour time blocks spread over 12 months to survey users of e-resources (primarily e-journals and databases). The 2013–2014 implementation, which for the purposes of this paper closed June 30, 2014, surveyed every 140th usage passing through the proxy server. The paper compares the two methods for reliability and validity of the results and ease of technical implementation and reports on the results of the recent survey, examining user demographics, time and date analysis, location of use, purpose of use, and collection development implications.

Further, this paper demonstrates how using business intelligence software for data analysis and visualizations to interact with the survey data in real time helped to:

- review and use live data throughout the year providing the ability to monitor the projected total results to make adjustments in real time (The survey frequency was increased from every 200th user to every 140th user to collect sufficient data to answer collection development research questions.);
- compare the distribution of sampled e-resources to all the usage of e-resources to judge the reliability of the sample;
- present a more informative visual display over SPSS and Excel graphics revealing relationships more easily and clearly; and
- collect survey data continuously, running the survey for the foreseeable future, and consider expanding the survey scope to include other resources.

Finally, the paper shows how data collected about users including status, academic affiliation, and purpose of use creates a deep picture of usage that can be combined with COUNTER data to give a more complete picture of electronic resource usage.

**MINES for Libraries**

As described on the MINES for Libraries website, [http://www.minesforlibraries.org/](http://www.minesforlibraries.org/), MINES stands for Measuring the Impact of Networked Electronic Services and is an online, transaction-based, intercept survey that collects data on the purpose of use of electronic resources and on the demographics of users, developed by Brinley Franklin and Terry Plum. MINES was adopted by the Association of Research Libraries as part of the New Measures toolkit in May 2003. It is a point-of-use web survey of three to five questions that integrates usage data about electronic resources such as digital collections, open access journals, pre-print and post-print servers, and institutional repositories, to give an inclusive picture of the library’s supported networked electronic resources.

In general, MINES for Libraries aims to:
• measure the value and impact of digital content;

• determine how specific user populations apply digital content to their work, based on demographic and purpose-of-use analyses;

• identify where library use originates in the networked environment and tailor services accordingly; and

• gather usage data about digital collections to justify increased funding for digital content and to make informed collection development decisions.

The roots of MINES are in indirect cost or facilities and administrative (F&A) cost studies focused on the library to help universities provide evidence for determining an accurate infrastructure support cost associated with sponsored research. These cost analysis library studies have been administered since the early 1980s by Brinley Franklin, although the web surveys have been in use since 2000. In addition to the web survey, the MINES methodology can also be used to assign a monetary value to a cost center in the academic library—such as a class of materials like e-journals or to specific vendors such as Elsevier’s ScienceDirect—to determine the portion of the cost center dedicated to the support of funded research, instruction, patient care, public service, and other activities by different classes of patrons of the library.

MINES employs a web-based user survey intercept methodology that delivers a short survey at the point of use of an e-journal, database article, or digital collection or service. There are two research designs or sampling techniques recommended by MINES, both of which result in a random sample of patron usage of networked electronic resources, the distributions of which can be applied with confidence to the user population. Note that in both cases the survey is a usage survey, not a user survey. MINES attempts to represent usage in the survey sample. The MINES protocol recommends that the questions on the survey are mandatory. Finally, even though the questions are mandatory, typically there are abandonments or patrons who back out of the survey. Depending on the implementation, these abandonments may be captured, and the respondents’ eventual responses, if any, may be recorded. We have found that 5,000 records collected over a year will answer most of the research questions posed by libraries that implemented MINES.

1. Random Moment Sampling

In the random moment sampling method for each month over a year either one or two two-hour survey sessions are randomly chosen for administering the web survey. The web survey then intercepts usage over that sampled period, querying the patron at the first usage of a surveyed resource, and repeatedly copying the values of the survey to a database at every subsequent use of surveyed resources by the patron during that two-hour period, adding a new time/date stamp and target URL or the URL of the surveyed resource. Currently, this sampling technique is employed for the cost analyses studies to determine F&A rates, and not for the ARL-supported MINES studies. The strengths of this sampling technique are as follows:

• The users are intercepted only 24 or 48 hours over a year, so the annoyance index is low.
• The number of records collected is usually sufficient to analyze usage by cost center.

• The data are collected at a common point, and therefore are commensurable (like COUNTER-compliant data) across diverse networked electronic resources.

• The survey collects data in a manner consistent with how patrons seem to use resources, that is, patrons often conduct a literature search intensely for a short period. The two-hour survey session tracks this burst of searching activity.

• The sample is random.

The weaknesses of this sampling technique are as follows:

• The intercept survey is intrusive.

• To log usage over a two-hour period the technologist must set up a session for the patron’s browser that will remember some token for the patron over the two-hour period. The session is usually established with browser cookies or a server-side session. Although there is no identifying information associated with the patron, it is important to this technique that the survey session knows the patron is still the patron.

• Depending on the intercept point, described in following paragraphs, the survey session may be more or less difficult to implement.

• Target URLs can be difficult to analyze. Some are dynamically generated, involve redirects, and may be hard to decipher by visual inspection as in the case of digital object identifiers (DOIs).

2. Systematic Sampling

With this sampling technique every Nth usage is sampled at some choke point or virtual gateway, such as OCLC’s EZproxy or an open URL link resolver, e.g., ExLibris SFX. The systematic sampling is an equal probability method within the ordered sampling frame and is often referred to as an every Nth sample. With this technique, an N is established and the starting point is randomly chosen. In libraries that have administered an every Nth sample, N has ranged from 1:500 to 1:140. Like the random moment sample, the every Nth data collection is conducted over a year to capture the different states of academic library usage: the academic year, the summer, and intersessions. The second Ontario Council of University Libraries (OCUL) study by ARL, OCUL, and the University of Toronto is a systematic sample for the 20 libraries involved with the study. ARL currently recommends this sampling method for MINES implementations.

The strengths of this sampling technique are as follows:

• The N can be changed to increase the number of records collected or to reduce the annoyance.
The most difficult part of the technical implementation for the random moment sample, that is, the creation of the session, is eliminated, so that the every Nth sample is technically easier to set up than the random moment sample.

The data are collected at a common point, and therefore are commensurable (like COUNTER compliant data) across diverse networked electronic resources.

The sample is random.

The weaknesses of the systematic sample are as follows:

- The intercept survey is intrusive.
- The every Nth sampling does not reflect patterns of heavy usage and light usage by individuals in the same manner as the two-hour session, although heavy users are likely to be surveyed more often over the year than light users.
- Depending on the intercept point, described in following paragraphs, the survey session may be more or less difficult to implement.
- Target URLs can be difficult to analyze. Some are dynamically generated, involve redirects, and may be hard to decipher by visual inspection as in the case of digital object identifiers (DOIs).

The intercept point for MINES should be some virtual gateway through which most users choosing a networked electronic resource must pass. Because the MINES survey methodology is based upon capturing the target URL or selected networked electronic resource, open access resources not included in the library’s electronic resource management system, bookmarks that do not include the proxy prefix, password-based alerting services, any services that depend on vendor passwords rather than some library mechanism, or e-books downloaded to e-readers are all problems for the MINES survey methodology. The intercept must be done locally at the library’s web services, unlike LibQUAL+® where the survey is accessible through the ARL platform.

Libraries have implemented MINES with different techniques and gateways, such as prepended Java, PHP or JavaScript redirects from a list of resources, a survey intercept at the proxy rewriter such as EZproxy, or a survey intercept at the open URL link resolvers such as Ex Libris SFX, Innovative Interfaces’s WebBridge LR, and ProQuest’s 360 Link. The SFX solution is described by Thomas and others. One version of an EZproxy solution has been written up by Reese. One of the advantages of the EZproxy implementation is that resources and services can be placed behind the EZproxy application, and therefore can be surveyed. For example, PubMed is often not behind the EZproxy server since it is a free resource. However, many libraries wish to include PubMed as a surveyed resource, in part because of its LinkOut feature, and it can be added to the EZproxied resources. In many cases the surveyed resource could be added as an open URL link resolver target also.
In another common networked-service technique that will increase the validity of the sample, the technology group at the university campus can push out the appropriate library access links (open URL link resolver) within Google Scholar to browsers on campus, thus increasing traffic to the link resolver and the proxy server if one is used, tightening up the web of possible survey points. The most comprehensive interception point is at the Internet service provider router for the university, and as radical as it may seem, this router-based solution has been implemented twice by one university and has been contemplated by several others. The router-based solution has the fewest limitations, but the other intercept points also work well, and collect reliable samples, as long as the limitations are understood.

**University of Massachusetts Amherst Environment and Implementation**

The University of Massachusetts (UMass) Amherst is a public research university offering undergraduate, graduate, and professional degrees. There are 28,518 undergraduate and graduate students, and 1,170 full-time instructional faculty. The UMass Amherst Libraries is the largest state-supported academic library in New England with more than 8 million items, spending $6 million on continuing e-resources.

The UMass Amherst Libraries implemented MINES twice, once using the two-hour random moment sample design in 2008–2009, and then the every Nth sample design in 2013–2014. The first survey ran for 12 months, from September 2008 to August 2009 with two two-hour survey periods per month. Figure 1 is a screen shot of the survey.

![Screen Shot of MINES for Libraries Survey at UMass Amherst](image)

**FIGURE 1. SCREEN SHOT OF MINES FOR LIBRARIES SURVEY AT UMASS AMHERST**
Because of technology limitations, the survey had to be manually turned on and off at the EZproxy server for each survey period. The help desk survey support changed from daytime to evening and the survey form was updated to reflect the changes each survey period. There were 4,396 completed surveys that were linked to URLs in the proxy log. The BioStatistics Consulting Group at the university ran the analysis on the data in SPSS, producing useful data but in the standard ASCII SPSS tabular format. Library staff then reworked these data in Excel, generating pie charts and cleaner tables. Figures 2a and 2b show how the visual presentation appeared.

<table>
<thead>
<tr>
<th>Question 3 - Location</th>
<th>Freq</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off Campus</td>
<td>2044</td>
<td>47</td>
</tr>
<tr>
<td>On campus but not in Du Bois or ISEL</td>
<td>1551</td>
<td>36</td>
</tr>
<tr>
<td>In the W.E.B. Du Bois Library</td>
<td>678</td>
<td>15</td>
</tr>
<tr>
<td>In ISEL</td>
<td>113</td>
<td>3</td>
</tr>
</tbody>
</table>

**FIGURE 2A. EXCEL TABLE OF RESPONSES TO MINES SURVEY QUESTION ABOUT LOCATION OF E-RESOURCE USE, 2008–2009**

The highlights from this study were that most of the users of networked electronic resources are not physically located in the UMass Amherst Libraries; there were a surprising number of undergraduates involved with research; and the libraries documented its contribution to the sponsored research endeavor as well as teaching and learning.

In 2013, the UMass Amherst Libraries implemented the MINES survey with the every Nth (N=140) systematic sample design. A Perl script with rules for presenting the survey was invoked using the EZproxy service banner redirect setting. The data was collected in a MySQL database, and live Tableau software connections were used to analyze and visualize the results. Figure 3 shows how the survey intercept works.
In this survey an abandonment reduction factor was included, making the survey more valid. Once the patron browser request was intercepted and the survey launched, the patron had five minutes to complete the survey. If the patron did not complete the survey in that time, then the survey timed out. A multi-step intercept calculation accounted for expired surveys and maintained the desired 140th user rate of return. The median time to complete the survey was 45 seconds (five minutes was the maximum allowed); the fastest survey was completed in 12 seconds. Figure 4 shows the survey completion time distribution in seconds. The overall response rate was 71%.
Tableau Software

Although Excel has its advocates for the analysis of the usage of electronic resources in libraries,18 we used Tableau Software19 for analysis and visualization in the second iteration of MINES at UMass. Tableau Software is a business intelligence software that can be used for data analysis and interactive data visualization. It is increasingly popular in analyzing library usage statistics.20 With Tableau, data results can be monitored in real time. Monitoring the results in real time proved useful because early in the study it was observed that an N of 1:200 was not collecting sufficient data to lead to reliable and valid results. With N at 1:200, 3,477 surveys would have been collected over the year, fewer that the 2009–2010 survey. By adjusting the N to 1:140 at the end of the first quarter roughly 5,035 records were collected.

Tableau compared the data collected at the Nth use with the data collected for all use at the proxy server to confirm that the distributions were indeed similar and the sample was a representative sample of the population of all EZproxy use. This validity analysis was done on the frequency distribution of web surveys by hour during the day compared to all EZproxy usage and the frequency of usage by month (see Figures 5a and 5b). As can be seen by inspection, the distributions are almost identical.

![Figure 5A. Hourly Mines Survey and All Proxy Traffic](image-url)
FIGURE 5B. MONTHLY MINES SURVEY AND ALL PROXY TRAFFIC

Using dashboards to display survey results visually helped the staff absorb and understand the survey results. The side-by-side presentation makes it easier to see relationships between responses within a single holistic view. The MINES User Group Dashboard (see Figure 6) shows the summary of all results. This broad overview provides the baseline for deeper analysis. For example, it is useful to see that 89% of overall use occurs outside the library buildings and that 53% of use was in support of teaching or classwork. Visualizing MINES data with Tableau increases utility because the data is easily filtered to answer a range of detailed questions posed by individual staff. A liaison librarian to engineering can see that 92% of graduate student use happens outside the libraries and that 45% of use is for thesis or dissertation work (see Figure 7). This kind of close analysis of the questions informs instruction, outreach, and support to constituents. The power comes not only from aggregate data or from a single conclusion but also from the ability to understand specific and integrated aspects of the data as needed for various purposes.
FIGURE 6. DASHBOARD SUMMARY OVERVIEW

FIGURE 7. LOCATION AND PURPOSE OF USE FILTERED BY GRADUATE STUDENTS IN ENGINEERING
Finally, one of the most interesting comparisons made using Tableau was to ingest COUNTER data for the Web of Knowledge into Tableau both for record views in the database and regular searches (see Figure 8).

![COUNTER Report for Web of Knowledge - Record Views and Regular Searches](image)

**FIGURE 8. COUNTER DATA FOR WEB OF KNOWLEDGE DISPLAYED BY TABLEAU**

The COUNTER data is useful but does not give the granularity of MINES data for the same resources. See Figure 9, which shows Web of Knowledge use by user type, purpose, school and college affiliation, the purpose of use, and the reason why the resource was selected.
At a glance it is easy to see that graduates students are the largest category of users and that the Web of Knowledge is used most heavily by patrons in the College of Natural Sciences. The table with shaded cells shows affiliation by purpose of use and the dark cells show that Natural Sciences uses the Web of Knowledge for both sponsored and non-sponsored research. The empty white cells are also easily seen. This level of granular detail, which can be filtered even further, is pertinent to designing library instruction programs, making collection development decisions, and marketing resources.

Implications

The UMass Amherst Libraries will continue to run the survey for the foreseeable future. The successful implementations, high response rate, lack of negative feedback, and the utility of the data have resulted in the decision to collect survey data continuously. The libraries will also explore expanding the survey scope to include other resources. Moving from an annual sample to continuous data collection provides current and ongoing data that is available to be analyzed alongside other data such as circulation,
interlibrary loan, building use, and network access; truly the culture of assessment as described by Lakos and Phipps.²¹

The relationship between MINES and COUNTER data may also be further explored. Specifically, it would be useful to determine how much electronic resource use is not captured by MINES. It may also be informative to examine the relationship between MINES, COUNTER, and vendor data more closely. It is challenging for some libraries to integrate separate COUNTER reports into a holistic picture of e-resource use, and one advantage to MINES is that all use is collected in a single data set. The MINES data set can be combined with COUNTER data in the Tableau environment for a deeper and more granular view of COUNTER data. Some libraries without access to Tableau might request of their electronic resource management (ERM) vendor the ability to import MINES data into their ERM to achieve a similar analysis. This deeper picture of patron activity is achieved anonymously without the need to track the path of individual patrons through various library and university systems and the associated implications for confidentiality, privacy, or ethics.

Conclusion

This paper contrasts the implementation of the two sampling designs for the ARL MINES for Libraries protocol at the same library, the University of Massachusetts in Amherst, discussing the advantages and disadvantages of each. The findings demonstrate why running the every Nth MINES study continuously is a good idea for libraries, and we described a valid and reliable implementation scheme using EZproxy. We show how using Tableau Software to analyze MINES data permits adjustments to data collection in real time, for example, by changing the frequency of N to answer new research questions. We compared the sampled data on certain variables to the population of data collected at the EZproxy server in the Starting Point URL (SPU) log files to show that the sample can be relied upon for valid inferences about the population. We illustrated how Tableau can present data relationships that can lead to decisions and actions by the library. We demonstrated how, with one vendor as an example, the joining of MINES data to COUNTER data can enhance the picture of how the resource is used and therefore, how the library might better serve patrons who consult that resource. Finally, we proposed that collecting MINES data continuously will lead to future service enhancements, especially if the data is imported into data visualization software, like Tableau, which makes the data easier to analyze, understand, and communicate. We anticipate that we will continue to collect valuable, actionable data to present a comprehensive picture of e-resource use to library and campus stakeholders, specifically informing collection development, instruction, support for research, marketing, and liaison work.

Endnotes


16. Ibid.


© 2016 Rachel Lewellen and Terry Plum

This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.
