Purdue University

Purdue e-Pubs

Charleston Library Conference

Usage Statistics: The Perks, Perils and Pitfalls

Christine M. Stamison
Swets, cstamison@us.swets.com

Nick Niemeyer
Annual Reviews, nniemeyer@annualreviews.org

Cory Tucker
University of Nevada, Las Vegas, cory.tucker@unlv.edu

Follow this and additional works at: https://docs.lib.purdue.edu/charleston

An indexed, print copy of the Proceedings is also available for purchase at:

http://www.thepress.purdue.edu/series/charleston.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information

Sciences. Find out more at: http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences.

USAGE STATISTICS: THE PERKS, PERILS AND PITFALLS

Christine M. Stamison (cstamison@us.swets.com) - *Senior Customer Relations Manager, Swets*
Nick Niemeyer (nniemeyer@annualreviews.org) - *Site License Manager, Annual Reviews*
Cory Tucker (cory.tucker@unlv.edu) - *Head of Collection Management, University of Nevada, Las Vegas*

Abstract

In these tough economic times usage statistics have become more important than ever as an essential tool for collection assessment and management decisions of electronic resources. While the advent of the COUNTER Code of Practice has made reporting usage more straightforward, it evident from reading the many library listservs that there is still considerable confusion on how to interpret these reports.

This presentation will first offer a short primer on the COUNTER Code and will cover some of the often misunderstood nuances that appear time after time in the listservs. The first presenter will engage the audience in a quiz of COUNTER: fact or fiction to test the audience's knowledge of the Code. Next, a representative from a small/medium sized academic publisher will advise on how best to work with publishers in order to get consistent reporting – whether COUNTER or not. While we often hear many complaints of how publishers are not producing consistent usage reporting, this publisher representative will enlighten the audience of many the pitfalls that can occur during the initial e-journal set-up that can skew usage statistics. Finally, a librarian will give real world examples on ways to use the data in a constructive and responsible manner.

Christine M. Stamison's Presentation

COUNTER Primer

While COUNTER (Counting Online Usage of Networked Electronic Resources) (www.projectcounter.org) was launched in 2002, there still appears to be considerable confusion on how to interpret usage statistics. The COUNTER Code of Practice was established to bring uniformity to the way statistics are collected and counted. This Code of Practice dictated the format, content and delivery specifications of usage statistics, and stipulated that reports be supplied free of charge.

To become COUNTER compliant a publisher/aggregator must pass a strict audit by a CPA, CA (chartered accountant) or other equivalent. The audit process is documented at http://www.projectcounter.org/r3/r3_E.doc. For a new provider, an independent audit is required within 6 months of compliance and annually thereafter. Among other things, the auditor is tasked with checking the report lay-out, file format and delivery against the COUNTER Code of Practice. Additionally, the auditor must be able to double-click in a quick succession on a link and have that register as one usage.

COUNTER Version 3.0 for Journals and Databases

Just this year (August 2008) version 3.0 of COUNTER for journals and databases was released. Therefore, only publishers adhering to version 3.0 of COUNTER can be designated as COUNTER compliant. As part of version 3.0 vendors must provide XML format reports so that statistics can be downloadable via SUSHI (Standard Usage Statistical Harvesting Initiative –

ANSI/NISO z39.93) in a pre-defined automated format. This allows libraries to more easily download and compare usage from multiple platforms and providers. It also allows the downloading of usage statistics into an Electronic Resources Management (ERM) system.

As many publishers began offering digitized back files of their titles, there arose a need to distinguish the usage of the current issues from the back file. Therefore, in version 3.0 all vendors/aggregators that provide journal archives as a separate purchases from the current journals must provide *either* Journal Report 1a: Number of Successful Full-text Article Requests from an Archive by Month and Journal (optional additional usage report in Release 2) *or* Journal Report 5: Number of Successful Full-Text Article Requests by Year-of-Publication and Journal. It is important to note that JR1 report <u>includes</u> the statistics for the back file therefore the librarian does not have to add up the statistics for JR1 and JR1a for a full view of the usage. The JR1a is merely a subset of JR1. This has been misinterpreted many times in the listservs.

With more investment in federated search and other automated search engines, anomalies in usage arose. As federated search engines scour through numerous databases and other files, hits from these go up even though the results may not be of interest to the searcher. Additionally, internet crawlers and robots have also posed a problem with inflating usage statistics. Crawlers and robots copy and index downloaded data for quicker retrieval and check links or verify HTML code – this potentially skews usage. In order to counter this inflation of usage, COUNTER version 3.0 now dictates that federated search and other automated search engines must be reported separately. Additionally, usage activity created from internet crawlers, robots, etc. must be excluded from COUNTER reports.

Finally, with the rise in the number of consortia and the amount of deals consortia manage, it became important to track the usage of these costly deals. COUNTER version 3.0 addresses this. With COUNTER 3.0 a vendor that supplies databases/packages to a consortium must offer usage reports. These reports are Consortium Report 1: Number of successful full-text journal article or book chapter requests by month (broken down by consortia member) and Consortium Report 2: Total searches by month and database (broken down by consortium member). The report format is XML only. The following reports should be provided by a journal or database provider in order to be COUNTER 3.0 compliant:

- **Journal Report 1 and 1a**
  Full-text article requests by Journal Title, Platform and Month
- **Journal Report 2**
  Turnaways by Month and Journal
- **Database Report 1**
  Searches and Sessions by Month and Database
- **Database Report 2**
  Turnaways by Month and Service
- **Database Report 3**
  Total Searches and Sessions by Month and Service
- **Consortia Report 1**
  Number of successful full-text journal article or book chapter requests by month
- **Consortia Report 2**
  Total searches by month and database

Ebooks also have their own usage reports through COUNTER. They are:

- **Book Report 1**
  Number of successful requests by month and title
- **Book Report 2**
  Number of successful **section** requests by month and title
- **Book Report 3**
  Turnaways by month and title
- **Book Report 4**
  Turnaways by month and service
- **Book Report 5**
  Total searches and sessions by month and title
- **Book Report 6**
  Total searches and sessions by month and service

As shown above, as new discovery techniques arose, the COUNTER code had to evolve to counteract any anomalies in usage. COUNTER version 3.0 has made significant strides in normalizing usage data that arising from new technologies. A journal or database provider should offer the following reports to be COUNTER compliant.

COUNTER Quiz

In order to gauge how knowledgeable the audience was on the COUNTER Code the speaker quizzed the audience on some of the nuances of the code.

Question 1: A turnaway report (JR2) shows:

    A. Number of times a user was turned away from an unsubscribed journal
    B. Number of times a user was turned away due to a faulty link
    C. Number of times a user was turned away due to exceeding the number of licensed users

Answer: C Number of times a user was turned away due to exceeding the number of licensed users

Time and time again librarians have posted incorrect information about the JR2 report on listservs. Many times the JR2 report has been interpreted as the number of times a user was turned away from an unsubscribed journal. In this case, the audience unanimously guessed correctly.

Question 2: In the COUNTER audit what is the space of time an auditor should be able to double click twice on a link without it being counted twice?

    A. 20 seconds for HTML, 25 seconds for pdf
    B. 10 seconds for HTML, 30 seconds for pdf
    C. 15 seconds for HTML, 20 seconds for pdf

Answer: B. 10 seconds for HTML, 30 seconds for pdf

This question was asked to gauge how much the audience knew about the auditing rules. Only a handful of the audience wagered a guess. The rest did not venture to guess. The audience was surprised to find out how this was gauged.

Question 3: If you have a subscription to an e-journal title directly from the publisher and through a database provider, in order to get a complete picture of usage you

       A. Can get the complete report from the publisher's website
       B. Can get the complete report from the database website
       C. Need to add usage together from both the publisher's and database
          providers' site

Answer: C. Need to add usage together from both the publisher's and database provider's site

In the past year there have been questions posted to listservs that indicated that users thought that all usage data would could directly from the publisher even if they had multiple ways of accessing content. The majority of this audience guessed correctly although a few were unsure.

The speaker then introduced the next presenter, Nick Niemeyer from Annual Reviews who covered how best to work with publishers to achieve consistent reporting.

Nick Niemeyer's Presentation

Introduction

Hello, I'm Nick Niemeyer, the Site License Manager at Annual Reviews. Annual Reviews is a non-profit publisher of 40 journals in various disciplines, so my remarks this afternoon are based upon journal usage. I imagine a lot can be carried over to e-books, but I don't have direct experience with that.

I want to talk about three aspects of journal usage statistics:
1. COUNTER Report pitfalls
2. Using usage statistics, some tips and warnings
3. Finally, something that isn't about actual usage at all, but is an extremely helpful report --- the Denials of Access To UNsubscribed Materials report.

1. COUNTER REPORT PITFALLS

Christine has given you the quick view of COUNTER Reports 1, 1a, and 2; and I'd like to highlight some potential problem areas.

A. **Journal Report 1 – Total Usage**

Journal Report 1 show you your TOTAL usage of all types; Journal Report 1a shows just the subset of Legacy usage. Do not add the two reports together, or you'll double-count Legacy usage.

As you see from the Journal Report 1 slide, COUNTER doesn't put the Grand Total Column where you expect to find it; it's not in the far right column. Instead, COUNTER puts the Grand Total in the third-from-the-rightmost column, followed by the HTML Download subset and then the PDF Download subset. This isn't a problem if you're using an automated data

harvester, such as SUSHI. But if you're cutting and pasting, be careful that you pick the correct column of data.

B. **Journal Report 1A – Legacy Usage**
This slide of Journal Report 1A shows the same unusual and unintuitive location for the Grand Total column. Just as with Journal Report 1, it's NOT at the far right where you expect to find it.

Another thing to keep in mind concerning Legacy usage is that different publishers have very different definitions of what constitutes Legacy material. For some publishers, Legacy might mean anything older than 1 year; for others, it might mean anything older than 5 years or 10 years. You need to ask the publisher before you do any serious analysis of this Legacy usage, to make sure you know what is and is not included.

C. **Journal Report 2 - Turnaways**
This report is my personal pet peeve, because I've had so many conversations with librarians who, understandably but mistakenly, take the report at its face title.

Once I phoned a librarian because her patrons had hundreds of access denials (which most librarians would think "turnaways" to be) for several of our journals. I reported that fact, and she said, "That's impossible. I just recently checked our Turnaways Report, and it was all zeroes."

Here's what I had to explain to that librarian. It is unfortunate that this report is called "Turnaways" because the report is extremely limited in its scope and far less useful than the general term "Turnaways" indicates. The report does not show turnaways in the general meaning of the term, but rather turnaways *due solely to exceeding the allowed number of simultaneous users*. Since many publishers, such as Annual Reviews, have no such limitation, the Turnaways Report will always show zero Turnaways for those publishers. There may actually have been thousands of turnaways, in the broad meaning of the term, but this report will never show them.

My advice is, unless you care specifically and exclusively about Turnaways due to exceeding the allowed number of simultaneous users, don't even look at this report.

D. **Account for all usage**
If you want to know the entire universe of your usage, remember that downloads from third party databases like ProQuest, EbscoHost, Academic Search Premier, etc., will seldom be reported by the publisher's own website. You'll need to get the statistics from those third party databases, as well as the publisher's statistics, to compute the true total of all usage ---- if that's what you want.

2. USING USAGE STATISTICS
OK, now that you've gotten the statistics, what do you do if there are some unpleasantly low usage figures? Here I won't talk about the librarian-to-librarian factors, since that's often been covered at Charleston, but rather the publisher-to-librarian factors.

A. **Are your IPs in place and are they correct?**
I know you'll find this hard to believe, but we have customers that have NEVER given us their IP range data. We have phoned and emailed, repeatedly over many years, to no avail. So, please, when a publisher tells you that there are no IPs registered on your account, respond to that message.

B. **Have you set up the Open URLs for your subscribed journals in your library systems?**

C. Have you publicized new acquisitions to your key stakeholders?

D. Have you avoided using multiple online accounts with your publisher?
Especially if you are placing separate subscription orders, as opposed to a site license order, it is important that the same account be used for all those orders. Otherwise, an order can become disassociated from the main online account, which is not a good thing because the chances for access problems increase as accounts multiply.

E. Have you changed agents from the prior year's orders?
If so, warn the publisher. Otherwise, unneeded new online accounts can be created, and the publisher may not realize that the IPs on your prior account X should be moved to a new account Y.

F. *Interpret* the cost-per-download figures
As past Charleston sessions have stressed, a bald cost-per-download figure must be interpreted to have meaning. Downloads can have very different values depending upon their end users. Do downloads from materials acquired for the hotshot Biology researcher who just snagged a $15 million NSF grant count the same to you as downloads from Newsweek Online? Only you can decide upon the final true value of your patrons' downloads.

3. NON-USAGE STATISTICS - THE DENIALS OF ACCESS REPORT
Up to this point, I've been talking about actual usage, that is, how many times your patrons download subscribed materials. Now it's time to look at the other side of the coin. Many publishers can supply you with a valuable tool that indicates unmet demand for UNsubscribed materials: the Denials of Access Report shows you the number of times your patrons tried to access UNsubscribed materials. Of course, this is not a COUNTER report, because it's not about usage. And not all publishers can supply this report, but some can.

This report has many uses. For example, it can substitute for a trial access period. I haven't given a trial in a very long time, because the libraries requesting trials all had their IPs registered with us, and I was able to tell them right away how often their patrons had been trying to access unsubscribed journals.

<Slide> Here's a summary view of the Denials of Access Report for a real institution, over a 3 ½ year time period. As you see, it breaks out the denials for each particular journal. That total figure of 31,000 denials demonstrates there's a lot of unmet patron demand at this institution.

<Slide> And here is a detailed view of the Denials of Access Report. It shows you both what year of a given journal your patrons tried to download, and also in what year they made their attempts.

The report can give you valuable insight into the existing demand for both legacy materials and current-content subscriptions. If you're thinking about acquiring a Legacy product, or perhaps adding some journals to your holdings, ask the publisher for this report in order to see the existing unmet demand. It can save you a lot of guesswork, and I'll bet that you'll see some surprises --- Like the medical school librarian who was astonished to learn about her patrons' demand for our unsubscribed Sociology journal, for instance.

SUMMARY AND QUIZ
So, we've looked at some issues related to COUNTER Usage Reports; we've gone over a few tips on using usage statistics; and we've seen the value of the Denials of Access Report, a non-usage report. That means it's QUIZ TIME!

QUESTION: The "COUNTER Report 2 – Turnaways" is useful for:
  a. Very little
  b. Driving librarians and publishers crazy
  c. Amazing your friends
  d. Justifying the $ spent on your MLIS
  e. Huh? Define "Turnaways"
ANSWER: e.

Cory Tucker's Presentation

UNLV Information

The University of Nevada, Las Vegas (UNLV) has over 20,000 full-time enrolled students and 800 faculty members. UNLV is considered a Carnegie Classification Research University (High). UNLV Libraries spends over 88% of the collections budget on electronic resources.

Evolution of Usage Statistics

UNLV has been collecting usage statistics for several years, so we have seen the evolution of usage statistics. Before COUNTER, the data was not reliable and the type of data varied from vendor to vendor. The data lacked consistency and standardization.
Once COUNTER was established and enforced, usage statistics became standardized and allowed libraries to compare "apples to apples." Standardization of usage statistics included terminology and type of data.

Although COUNTER has been extremely helpful for libraries, there are still some issues with vendor reporting. There is one of the large vendors where we have discovered a potential issue. The vendor in question has both Abstracts and Indexes (A&I) and Full-Text (F-T) databases. The statistics from the vendor show full-text downloads/views for the Abstract and Indexes. This reporting shows where the Abstract and Index database provided a link to a Full-Text database. Due to the fact that this type of usage shows up in the statistics for the Abstract and Index database, what is the impact on the statistics for the Full-Text database? There is a potential issue with "double-dipping."

So, what does the future hold for COUNTER and usage statistics? There are still issues to be resolved. One issue is with specific resource types like e-books. Some e-book vendors provide LC class, while others do not. Also, there are issues with journal titles. Some have stats only under new journal title, others have statistics for both the new and old titles listed and include stats for both.

Why UNLV Collects Usage Statistics

UNLV Libraries has been collecting usage statistics for over ten years. UNLV also has a President who is very data-driven. We provide him with semi-annual usage statistics on databases and journals. In addition to sending usage statistics to the University President, the Libraries use statistics to:

  • Begin the investigative process
  • Justify the collections budget

- Help improve marketing and promotion practices
- Provide feedback to vendors

For UNLV Libraries, usage statistics are just the starting point. Usage statistics allows the Libraries to:

- to cancel subscriptions
- to upgrade existing resources
- to aid in strategic planning
- to obtain "end of year" money from University Administration
- to compliment existing subscriptions with addition of backfile purchases
- to help make decisions in Add One/Drop One
- to aid in collection assessment

Currently, UNLV Libraries obtains usage statistics from 96% of our vendors. Table I shows the Libraries' vendor information:

Fiscal Year	ER Reporting Stats	ER Not Reporting Stats	Total ERs
2000	62	13	75
2001	102	19	121
2002	102	21	123
2003	102	21	123
2004	134	2	136
2005	200	7	207
2006	247	8	255
2007	286	12	298

UNLV Libraries has a "home-grown" system. The Libraries have investigated other products and found that it is more cost-effective to go this route. Harvesting statistics is very time consuming and statistics should be harvested in the most efficient manner. UNLV Libraries has a unique way to harvest the data:

- Harvest statistics each quarter
- Harvest statistics after the 15th of the month
- Student worker harvests statistics

- Support staff inputs data & audits/solves problems

Support staff in collection management does audit the usage data we collect to make certain there are no issues.

Usage Statistics Harvesting Process

A vendor spreadsheet was create to allow the department to have the information easily available for the student to harvest the usage data. The vendor spreadsheet contains a variety of information including:

- Vendor name
- Titles of electronic resources
- Distribution
- URL
- Notes
- Login
- Password
- Date record is updated

Once the usage data is harvested from the vendors, we aggregate the raw data into a spreadsheet and calculate data including:

- Vendor
- Title of electronic resource
- Monthly searches & full text views
- Total Searches & full text views
- Annual cost of product
- Cost per search & full text view

Evaluation and Usage Statistics

Harvesting data is time consuming and it is just the beginning of the evaluation process. The big question is "now that we have the data, how are we going to organize it and what are we going to do with it?" UNLV Libraries decided that we needed to enhance the data by incorporate statistics from *Serials Solutions* and queries from the Libraries' link resolver. In addition, the Libraries wanted to include inflation calculations, including percentage increase/decrease and calculate trends in usage, also by percentage increase/decrease. Finally, for all of the Libraries' journals, the data for all access points was entered into one spreadsheet.

After all of the data was gathered, the Libraries decided to organize the electronic resources into "like" or similar groupings. The groups of electronic resources are:

- Electronic journal subscriptions
- Index & Abstract databases
- Full-text Aggregator databases
- Other Full-text databases
- Electronic journal packages
- ebook subscriptions

Once this was finished, a benchmark was calculated for each group. Another decision was to use three years of usage data for evaluation. This would allow time for the resource to be marketed and/or discovered by users.

Evaluating E-Resources

Although usage statistics are very important in the evaluation process for electronic resources, they are only a starting point. When usage is low, you need to attempt to discover why it is low? Has the database been marketed to users? Are there issues with access? Just because a journal or database has low use does not mean the resource should be cancelled. A library needs a variety of information to make best decision. Because of this fact, the most effective way to evaluate electronic resources is to combine usage with other evaluation techniques:

- Focus groups and surveys (faculty and students)
- Grant awards
- University-related information (demographics, programs, strategic planning documents, etc)
- Faculty publications – where are they publishing and what are they citing
- Overlap/Duplication

Once all of this information is harvested and analyzed, it needs to be passed along to librarians to help with collection development decisions. The key points when working with librarians who are selectors is that the collection management/development department needs to be in constant communication with librarians. The usage data needs to be placed in a format that can be easily understood and used by librarians. In addition, the usage data should be easily accessible. Although this data can be very useful for subject librarians, I do often wonder how many librarians actually use the data in their collection development decision making process and what type of value does the data hold for the selector.