

3-2013

A Hybrid Model for Managing Standard Usage Data: Principles for E-resource Statistics Workflows

Amy Fry

Bowling Green State University - Main Campus, afry@bgsu.edu

Follow this and additional works at: https://scholarworks.bgsu.edu/ul_pub



Part of the [Library and Information Science Commons](#)

Repository Citation

Fry, Amy, "A Hybrid Model for Managing Standard Usage Data: Principles for E-resource Statistics Workflows" (2013). *University Libraries Faculty Publications*. 8.

https://scholarworks.bgsu.edu/ul_pub/8

This Article is brought to you for free and open access by the University Libraries at ScholarWorks@BGSU. It has been accepted for inclusion in University Libraries Faculty Publications by an authorized administrator of ScholarWorks@BGSU.

1. Introduction

The migration from print to electronic collections over the last two decades has created a new environment for gathering and assessing collection use. Circulation and reshelving statistics tabulated in-house for journal and index volumes have been replaced by sophisticated, automatic monthly recordings of logins, searches, and full-text downloads provided by database and e-journal vendors. Because libraries often subscribe to thousands of titles from hundreds of sources, they have become awash in a flood of usage data that has proven difficult to collect, manage and assess, despite its value.

An early assessment and explication of the problems with usage data led to the emergence of standards, most notably the COUNTER standard, released in 2003. Parallel efforts to create systems to accommodate standards-based statistics transformed homegrown methods for counting usage and collating usage reports into hybrid methods using both locally developed tools and third-party tools (including ERMs and assessment products) that could take advantage of these standards.

This article will outline principles for creating a hybrid workflow for e-resources usage statistics, using the experience of Bowling Green State University's Jerome Library as a case study.

2. Early Problems with E-Resources Usage Data

Challenges related to e-resources usage data have long centered on four main issues: data collection, data transfer/manipulation, incomparability, and integration (Fons & Jewell, 2007). In 2000, Luther was commissioned by the Council on Library and Information Resources to review the availability of e-resources usage statistics and the issues surrounding them. She interviewed librarians and vendors and identified issues of concern to both groups, including lack of comparable data, lack of context for the data provided (use context and content context), and partial availability (data not offered for every resource), among others (Luther, 2001).

Also in 2000, participants in the ARL E-Metrics study made the following complaints about vendor usage statistics: they were not detailed enough, were inconsistent (differing from vendor to vendor and lacking a clear explanation of what was measured), were not comparable (measured different things), or were not available. The study concluded that libraries would not invest much in gathering statistics until they were consistent, and recommended the implementation of a standardized format for reports with "common data elements and arranged in a predetermined, agreed-upon order" (Shim & McClure, 2002, p. 508).

3. Emergence of standards

The International Coalition of Library Consortia (ICOLC) developed the first guidelines for reporting e-resources usage statistics. Released in 1998, they defined the elements that should be counted, including logons, queries, and items examined (Blecic, Fiscella, & Wiberley, 2001).

Project COUNTER (Counting Online Usage of NeTworked Electronic Resources) went a step further by specifying not only which elements should be included (searches, sessions and full-text downloads) but *how* vendors should measure and report them. With COUNTER, the industry finally established something that a number of studies had recommended – a standard for data validity that would ensure that statistics could be compared across different vendors and platforms. Now in its fourth release, COUNTER includes data elements, protocols, standard reports in standard formats, and an auditing requirement (Shepherd, 2006). It has evolved according to its original mandate from librarians, who, in a 2002 survey, clearly stated that they wanted the project to focus on the reliable production of basic

reports downloadable monthly from a secure website in a format compatible with Microsoft Excel (Shepherd, 2004).

In 2007, Fons and Jewell stated that, though they believed wide adherence to COUNTER would improve libraries' ability to work with usage statistics, "without other improvements these tasks will likely still be overwhelming" (p. 156). Indeed, despite precise formatting guidelines and the auditing requirement, inaccuracies in the format of COUNTER reports have long existed which can complicate working with them (Baker & Read, 2008; Blecic, Fiscella, & Wiberley Jr., 2007; Fons & Jewell, 2007).

Another lingering difficulty is the process of gathering these reports, or the expense of paying a service provider to do so. "The major reason [usage statistics tasks are overwhelming for many libraries] is that the process of gathering the usage data is typically quite time-consuming; each source of data is likely to have its own access idiosyncrasies...as well as require multiple steps following login," Fons and Jewell wrote (p. 156). Many respondents to Fleming-May and Grogg's 2010 survey indicated that they gathered statistics manually and expressed frustration with this process (2010b).

The SUSHI (Standardized Usage Statistics Harvesting Initiative) protocol was developed in 2005 and 2006 in an attempt to address this difficulty (Fons & Jewell, 2007). SUSHI enables the automatic transfer of COUNTER data between systems and was added to the COUNTER Code of Practice in 2008 (National Information Standards Organization, 2008). However, to date, SUSHI has failed to live up to its promise, largely because of the lack of systems available to take advantage of the protocol and the ongoing irregularity of vendors' applications of the COUNTER standard (Fleming-May & Grogg, 2010b, p. 30). In 2011, Collins and Grogg found that "librarians' most common complaint [about electronic resource management systems (ERMs)]...is the failure of ERM systems to implement the Standardized Usage Statistics Harvesting Initiative (SUSHI) standard" (2011, p. 24).

4. Early Statistics Workflows: Homegrown

Because early e-resources usage statistics were so irregular, early projects for counting local e-resource usage often dispensed with vendor-provided statistics, using, instead, homegrown systems to capture statistics that librarians felt were more reliable, comparable and complete. In 1996, Hyland and Wright combined non-intrusive (automatic) and intrusive (user-survey) methods to generate local usage data for CD-ROM databases at the University of Wollongong. The researchers concluded that usage data should be collected over months and years to provide a truer picture of use. Blecic, Fiscella and Wiberley reached a similar conclusion after analyzing e-resources usage data at the University of Chicago in 2001.

In 2001, Duy and Vaughan compared usage of electronic resources tracked from North Carolina State University's website to vendor-provided data, performing correlation tests on session counts for 28 different resources. While patterns and values in the two data sets were similar and the authors thought it might be valuable for libraries to use locally-generated usage data to check the accuracy of vendor data, they also concluded that there was no real substitute for vendor-provided usage data and called for more standardization (Duy & Vaughan, 2003).

A third homegrown project for recording e-resource usage statistics was carried out at the Texas State Library and Archives Commission from 2001 to 2003. The consortium automated the process of tracking usage data for its metasearch application by developing a log analysis tool to record searches, sessions and full-text downloads. The tool provided the consortium with data that could be compared across resources and institutions (Moen, Oguz, & McClure, 2004).

A 1999 project at the University of Louisville compellingly illustrates the pitfalls of generating usage statistics for e-resources in-house. Librarians there had been collecting and compiling vendor reports where available, but wanted a more consistent, automated method of measurement that could apply to all of their resources. Therefore they used ICOLC guidelines to implement a common gateway interface tracking program that automatically recorded accesses of journal titles from the library's e-journal lists. This data was then extracted daily into an Access database for reporting (Liu & Cox, 2002).

While this provided the information the library wanted, it required users to connect to journals from a list generated by the library's catalog in order to record the use. It also recorded only one type of use – number of accesses, and required daily extraction. Such a system lacks the agility needed to provide a true picture of patron behavior by not being able to record serendipitous discoveries of content and only recording one kind of use, and is also burdensome on the library both in terms of frequent, labor-intensive processes and constant system updating. Drawbacks such as these convinced libraries, as standards emerged, to commit to wrestling with vendor data, both standard and non-standard, and develop workflows that could take advantage of this data, with or without third-party tools to help them.

5. Migration to Hybrid Workflows

The literature shows that, in the years since COUNTER was released, libraries have moved away from trying to count use themselves and towards using vendor-supplied standards-based usage reports. They have also moved away from trying to deal with these reports through completely homegrown methods (workflows based on Excel spreadsheets and Access databases) to a hybrid of homegrown methods and third-party products.¹

In 2004, Bordeaux and Kraemer described statistics workflows that did not use any third-party e-resources management software or services. Kraemer had developed a relational database at the Medical College of Wisconsin that could combine current and historical COUNTER usage data with cost information, while Bordeaux described a workflow based on Excel spreadsheets in use at Binghamton University (Bordeaux, Kraemer, & Sullenger, 2005). In 2009, Matthews reported that the Virtual Library of Virginia Consortium was using, and committed to maintaining, a locally-developed workflow for collecting and processing COUNTER-compliant statistics. Also in 2009, Tucker reported that the University of Nevada Las Vegas was using an Excel spreadsheet to manage their usage statistics locally because they could obtain statistics for more of the library's e-resources that way than by using a third-party service.

However, in another 2009 article, Walker, after describing the Wichita State University Libraries' locally-developed, Excel-based stats workflow, admitted that, in the future, that institution planned to migrate to a hybrid model in which it would use an ERM to help manage its statistics. Why? Though Walker maintained that spreadsheets were the simplest way to deal with vendor usage statistics (p. 248), the workflow she described was highly complex, labor-intensive, and potentially prone to human error. Staff at WSU maintained separate spreadsheets to record and report fiscal-year usage and cost-per-use for databases and e-journals, many values were keyed by hand, and the same data was often kept in multiple places. While Walker said it was advantageous to have all data under local control, she also

¹ These include commercial and open-source electronic resources management (ERM) systems such as Innovative Interface's ERM and Serials Solutions' 360 Resource Manager (for more, see Collins, 2008) and assessment products/stats-gathering services such as ScholarlyStats (now part of Swetswise Selection Support) and 360 Counter (for more, see Paynter, 2009).

admitted it was time-consuming to download, manipulate and enter information into these spreadsheets, and even more time-consuming to extract the desired information. Therefore, she planned to migrate to a hybrid process.

In 2010, librarians from Bowdoin and the University of North Carolina Wilmington (UNCW) both reported using a combination of third-party software and locally-developed practices to manage usage statistics. At Bowdoin, a homegrown system for housing statistics was briefly replaced with a third-party assessment product (Serials Solutions' 360 Counter) before the library decided that "moving everything into an ERM system and developing a homegrown hybrid" of a vendor-supplied ERM and an open-source system would be the best solution (McQuillan, Fattig, Kemp, Stamison, & England, 2010, p. 112). UNCW's hybrid process included subscribing to a stats downloading service (Scholarly Stats) in addition to manually downloading vendor statistics, uploading COUNTER Journal 1 (JR1) reports to an ERM, using the ERM to store management information for gathering stats, and manually compiling statistics into a spreadsheet for reporting (McQuillan et al., 2010). Hulbert, Roach and Julian were also using a hybrid model for managing statistics at the University of St. Thomas in 2011. There, they used an ERM (360 Resource Manager) and an assessment product (360 Counter) in conjunction with a locally-created Access database to generate usage reports (Hulbert, Roach, & Julian, 2011).

Increasing availability of COUNTER statistics supports the implementation of software to work with this data, but the lingering presence of non-COUNTER-compliant statistics demands the maintenance of manual processes – hence the attractiveness and necessity of hybrid workflows like the ones these authors describe. A look at the evolution of statistics for Bowling Green State University's e-resources subscriptions illustrates this. In 2011-2012, the library offered 177 subscription database products, 27 more than in 2007-2008. In both years, the library was able to obtain usage statistics for 93% of its subscriptions, but the ratio of COUNTER-compliant to non-COUNTER-compliant resources changed dramatically over this time period. In 2007-2008, the library was able to obtain COUNTER-compliant statistics for only 39% of its database subscriptions, but in 2011-12, COUNTER-compliant data was available for two-thirds of its resources. This trend means it makes more and more sense for libraries to implement statistics processes that take advantage of e-resources management tools that incorporate the COUNTER standard, even though plenty of resources remain outside of this standard.

The fact that libraries are not yet able to rely on a single broadly-available standards-based system or service to manage statistics, however, illustrates the ongoing difficulties libraries have with statistics. The need to bring together data from a range of vendors in a wide range of compliance with standards and services, the need to maintain legacy systems to retain non-standard historical statistics, and the need to combine use data with pricing if cost-per-use is desired means, despite their cost, third-party software and services only partially fulfill libraries' needs, a fact that has led some libraries to reject these products or discontinue their use.²

6. Establishing a Hybrid Statistics Workflow

Moen, Oguz and McClure recommend libraries address the following ten items when establishing an e-resources usage statistics workflow (2004, p. 420):

1. overall responsibility for data management
2. specific responsibility for collecting data

² One respondent to Fleming-May and Grogg's survey reported discontinuing ScholarlyStats because she felt it wasn't helpful enough: "Once we set up our admin accounts, it doesn't take an undue amount of staff time to retrieve [our usage statistics]," she wrote (Fleming-May & Grogg, 2010b, p. 31).

3. designing and maintaining a management information system
4. entering data into this system
5. verifying data with vendors
6. analyzing data
7. reporting data
8. knowing about the budget available to support this activity
9. knowing about the audiences for reporting
10. ensuring staff have the necessary skills to complete this process

Portland State University's usage statistics workflow document outlines that institution's hybrid workflow (2011). It includes three workflow and three data elements necessary for working with usage statistics at that library.

Workflow elements:

1. Record management (includes recording administrative information and contacting vendors for problem-solving)
2. In-house statistics gathering (includes manually gathering and recording statistics from vendors for those resources that require it)
3. 360 Counter administration and maintenance (outlines workflows for both COUNTER-compliant and non-COUNTER-compliant resources)

Data elements:

1. Historical usage data (saved in one archive)
2. List of resources included in 360 Counter
3. Lists of resources by type of statistics available (COUNTER-compliant and non-COUNTER-compliant)

The 2004 DLF ERMI (Digital Libraries Federation Electronic Resources Management Initiative) report includes a flowchart for e-resources workflow. It addresses usage statistics only once: as an activity within the activation/installation stage of the e-resources management lifecycle in which the library determines whether usage data is captured locally, from the vendor, or both. A footnote to the next step in the resource's lifecycle, "routine product maintenance," indicates that this step includes regularly collecting usage statistics (Parker et al., 2004, p. B-8). However, the report does not provide guidance on how best to pursue that work.

Phase II of ERMI, launched in 2006, proposed to deliver "a statement of functional requirements for vendor/publisher-supplied usage data harvesting, management, and reporting"(ERMI 2 Steering Group, 2006). However, though the project provided for the launch of SUSHI, it did not address workflow, instead completing a standards gap analysis (Anderson, 2009) which eventually identified workflow management as a gap and provided links to workflow documents from a number of different institutions to foster discussion across the library community (Jewell et al., 2012, para. 24).

To fill the workflow management standards gap, the TERMS (Techniques for Electronic Resource Management) project was launched by Jill Emery and Graham Stone in 2011. It defines six stages in the e-resources life cycle and is attempting to create, with the input of the library community, workflows for each. Libraries are encouraged to test the process of obtaining statistics in the implementation stage (stage three), and statistics are mentioned as an evaluation tool in stage four (ongoing evaluation and access). In stage five, annual review, the TERMS project gives some advice on collecting and reporting statistics, but does not outline a workflow (Emery & Stone, 2011).

It is clear that an effective statistics workflow must, for now, be a hybrid model, taking advantage of standards, available software and services (including automation), and local institutional knowledge. It is clear, too, that managing database usage statistics is an evolving practice, and any workflow implemented must also be organized for change: changes in software, changes in resources, changes in staffing, and changes (or evolutions) in standards. When establishing or revising a workflow for statistics, these principles should be applied to each part of the process.

Despite the limitations of any particular tool used in a hybrid model, establishing such a model is both feasible and desirable in today's e-resources management environment. It not only incorporates existing standards and uses technology to promote efficiency and accuracy, it also integrates usage statistics at all stages in the lifecycle of each e-resource. In 2011, the Jerome Library at Bowling Green State University (BGSU) migrated a statistics workflow that had been dependent on local files and practices to a hybrid model that took fuller advantage of existing standards and third-party software and services (see Figures 1 and 1a). While this workflow is specific to BGSU's software, collections and staff, it can be adapted to any library's. Work at each stage is defined as either traditional (e.g., completed by manual processes) or automated (completed within or with the aid of a standards-based third-party tool for e-resources management, such as an ERM). Each task is further defined as being completed by librarians, staff, or student workers. The model allows for fluid movement between traditional and developing methods of managing statistics as software, staff and materials are added or change.

In addition to addressing the items of consequence listed in Moen et al., the workflow and data elements recorded by Emery, and the signposts from ERMI and TERMS, BGSU decided to apply four additional principles in developing this hybrid process:

1. controlling data
2. simplifying work
3. redistributing work (involving staff at varying levels)
4. making use of existing standards and software

The resulting workflow provides a more accurate and complete picture of the library's e-resource usage in a way that is both more efficient and more agile, taking advantage of existing strengths while being adaptable for future developments.

6.1 Institutional profile

BGSU has been collecting and reporting e-resource usage statistics for its database and e-journal subscriptions since approximately 2003. Through 2010, acquisitions and serials staff shared responsibility for collecting database statistics, and reports were typically compiled by the Collections and/or Serials Librarian and reviewed annually at a meeting of reference and collection development librarians for possible database subscription changes. Statistics for consortially-purchased e-journals were reviewed annually when processing subscription renewals. Statistics for all e-resources were also gathered and reviewed as requested by subject librarians and for national reporting.

The Collections & Technical Services Department underwent significant organizational and staffing changes in 2010 and 2011, ultimately experiencing a 40% reduction in staff. It was acutely necessary to both reassign work in order to cover vacancies as well as create more efficient work processes wherever possible. The process of collecting and reporting statistics was one place where the department felt reorganization and automation could save time, as well as result in a more effective process.

The library was especially committed to making more effective use of its electronic resources management module, Innovative Interfaces' ERM. The module had been purchased and implemented in

2005 chiefly to manage the databases A-Z list and databases-by-subject pages on the library's website, as well as to help integrate the e-resources acquisitions process with that for other library materials. But some functionality of the software, particularly its ability to help with statistics gathering, archiving, and reporting, had never been implemented. As the process for streamlining the statistics workflow began in 2010, making better use of this tool without adding new systems that duplicated its functionality was a primary goal of the project.

6.2 Controlling data

Because responsibility for gathering e-resources usage statistics pre-dated the implementation of the ERM, and because it had been shared by different technical services units and a number of different staff members, the library had a bewildering array of incomplete and incorrect information related to statistics. Logins, passwords and instructions had been saved in a variety of spreadsheets on shared and personal drives and irregularly updated. Statistics files – originals, copies, and incomplete files – were scattered across departmental shared drives with no conventional naming standards. The resulting duplicative information made it extremely difficult to find information when it was needed, compile saved historical reports, guarantee that information remained accurate, and ensure that information was available to the people who needed it. Therefore, the first part of the project had to be controlling data – making sure each piece of information was available and correct and that it was stored in one place where it could be both secure and appropriately available.

First, and most importantly, the library transferred all administrative logins, passwords, and urls for accessing administrative and statistics modules for each database from spreadsheets into the ERM, using the "administration" field in each resource record. This default variable field has a security authorization separate from that for viewing and editing the rest of the resource record, so it is easy to control exactly who can see and edit each database's administrative login. Using the ERM for this information not only makes it available to whoever needs it, it also makes it more difficult for any one person to export, duplicate or change it.

The library also repurposed a field in the ERM to contain information about the availability and retrieval method of usage statistics. This fulfills the third data element in the Portland State University workflow (lists of resources by type of statistics available) without the need to actually maintain (and update) any separate lists: a query in the Innovative ILS will generate up-to-date lists for gathering statistics each year.

Further, the library edited the "usage statistics" variable field, removing administrative login information (now housed in the secure administration field) and replacing any existing instructions related to gathering statistics with a link to an external wiki page containing this information. Two wikis created by staff in 2011 – one for databases and one for e-journals – cover over 100 database and e-journal platforms on as many individual wiki pages (Author & Staff, 2011a; Author & Staff, 2011b). Each page includes a list of the resources on that platform, which reports need to be saved, and how to download them, including how to save the files, what to name them, and what to do with the saved files for reporting. Because many different databases can share one set of retrieval information, these wikis help the library avoid duplication. Instead of copying downloading instructions to the record for each resource on a given platform (which can include dozens of databases) and then updating all of these when the downloading instructions change, they are instead recorded once on an easily-edited web-accessible wiki page which is linked to the resource records for each database on that platform via the usage statistics field. Having instructions available means that downloading files can be assigned to anyone and time isn't repeatedly wasted figuring out the steps to follow.

Wiki software is ideal for this kind of information because it is secure, searchable and easy to update. This is crucial, because this information changes frequently: vendors consolidate, resources change platforms, companies update statistics and administrative websites, and databases themselves are added and canceled. The wikis are publicly available for viewing both so they can be shared with any staff member or student worker assigned to gather statistics and so they can be used as a resource by any library who also collects usage statistics for any of the database platforms included.

Finally, the library consolidated and restructured how statistics files are saved after being downloaded, again paying attention to securing the files and avoiding duplicate information. All reports are saved in two folders on a shared drive, access to which is governed by an individual's login to the campus network. The files are organized by platform and database name. Each statistics file itself includes the name of the database (making it searchable) as well as the report type. Wherever possible, the statistics for all years of a given resource are saved on different worksheets in the same Excel file, making it easy to compile historical statistics for review.

6.3 Reorganizing work

Once this data was under control, the library reviewed its statistics workflow for sustainability, trying whenever possible to break the work down into its smallest parts, simplify and standardize it, and integrate it with other workflows surrounding the lifecycle of each e-resource.

Integrating statistics into the overall e-resources management process was key. Obtaining administrative logins, filling in the three ERM fields used for statistics, adding resources to the appropriate wiki pages, and creating stats folders on the shared drive is part of the e-resources acquisitions process. Removing this information is part of the deacquisition process. Migrating it when platforms change is part of the updating process. The library uses checklists which follow detailed policy workflows for each of these functional parts of the lifecycle of an e-resource, and all checklists and policies include steps related to statistics. This simplified the process of gathering statistics and helped the library move to an annual rather than a monthly workflow for this work.

Historically, many libraries have collected usage data monthly, even though it is more time- and labor-intensive to do so. Binghamton University was spending approximately ten hours a month on statistics in 2004 (Bordeaux et al., 2005, p. 298). In 2009, Wichita State University was spending four to eight hours a month gathering statistics for databases and another eight hours entering this data into a spreadsheet (Walker, 2009, p. 239). The University of North Carolina Wilmington also gathered statistics monthly in 2009 (McQuillan et al., 2010, p. 112) while, in 2007, UNLV reported doing this work quarterly (Tucker, 2009, p. 51).

A survey conducted in 2010, however, shows that workflows were changing, most likely in response to evolving standards. While a slight majority of respondents still reported collecting stats monthly (51) or quarterly (11), almost as many had begun to do this work annually (45) or twice a year (16) (Fleming-May & Grogg, 2010b, p. 28). COUNTER compliance supports moving to an annual workflow, as each COUNTER report includes only data for that calendar year. If data is required internally before it has been downloaded, many libraries may find using estimates to be adequate. A small sample of database usage statistics at BGSU shows that fiscal-year estimates created by doubling fall usage (July through December) are often conservative estimates of the actual usage numbers (see Table 1).

Therefore, BGSU decided it would be most efficient to collect and collate database statistics each spring and e-journal statistics each fall. Such a workflow guarantees the availability of data for annual review but is also more efficient and makes the best use of available student labor (concentrated in spring and fall) for both cycles.

In the past, staff members collecting statistics monthly would also finesse the reports and rebuild them into spreadsheets that contained monthly search counts saved by fiscal year. This not only added complexity to this process, but dismantled any standard formatting. A hybrid process that makes use of third-party tools requires leaving COUNTER reports intact. Therefore, the library also decided to immediately stop editing COUNTER-compliant statistics files and start saving them by calendar year, as downloaded. The library also saves statistics for non-COUNTER-compliant resources as monthly counts by calendar year whenever possible.

6.4 Redistributing labor

Because the work was reorganized, staff at different levels (professional, paraprofessional and student workers) can be responsible for parts of the process for every resource. At BGSU, the e-resources librarian is responsible for the overall management of the statistics workflow, managing work related to statistics at each stage in the life cycle of each resource, making sure statistics are available when needed, and compiling reports. A student worker gathers statistics files from the vendor websites according to the instructions in the wikis, converts JR1 files from CSV to XML and uploads these into the ERM. Once staff are available for e-resources management (currently, all full-time staff are either assigned to either serials or acquisitions), a staff member can be assigned to maintain the wikis and update resource records when information about collecting statistics changes. This work is currently done by the e-resources librarian.

6.5 Reporting

To mitigate the difficulty of working with vendor usage reports, many libraries have limited the information they collate for reporting, choosing just one of the three main COUNTER measures to include for all resources. There has, however, been some disagreement over time about which measure (searches, sessions or full text downloads) to use. In 2005, Coombs wrote that libraries should look at session counts, because searches can be impacted by the skill level of the user, the usability of the resource, the availability of full text, and the resource's reliance on specialized vocabulary. Thus, she concluded, "it is inaccurate to compare database usage utilizing number of searches" (p. 599). This contradicts Bleicic, Fiscella and Wiberley, who wrote in 2001 that only looking at sessions could be inaccurate, because sessions with no searches could indicate a mistaken use of a resource (p. 446). The majority of respondents to Baker and Read's survey agreed, indicating that, for non-full-text databases, searches were a more useful indicator of use than sessions (2008). However, in 2007, Bleicic, Fiscella and Wiberley decided that federated searching had made both searches and sessions less meaningful counts that they had been in the past (p. 42).

Bordeaux reported only tracking searches at Binghamton University (p. 298), while UNLV tracked both searches and full text views (Tucker, 2009, p. 51). Only Hulbert, Roach and Julian reported collecting all three uses (2011, p. 159).

At BGSU, the library had historically reported only searches for its databases, as this was the measure of use most widely available. However, expanded openURL linking and discovery platforms are rendering this approach inadequate. Users might turn to a general database like Academic Search Complete or a discovery layer like Summon for most of their searching, but connect to a different database to

download an article. In fact, Way showed that implementation of a discovery layer may cause database searching to drop significantly, but full text downloads to grow at an even greater rate (2010). Therefore, not examining sessions and full text downloads at the database level could cause the library to miss growing use of a resource. Likewise, as relevancy ranking in research databases improves and becomes the default setting (certain vendors, like EBSCO and LexisNexis, only recently made this change), users may be completing fewer searches, even as use of a database (represented in session counts) remains steady or even rises. Therefore, BGSU is now committed to reporting searches, sessions *and* full text downloads, where available, for all COUNTER-compliant databases. For non-COUNTER-compliant resources, library chooses data elements that are comparable to searches, sessions and downloads and ignore most other counts (like “hits,” “pageviews,” or numbers of search results).

Database reporting, however, remains a challenge. Most assessment tools and ERMs, including the Innovative ERM, are still designed to consolidate and report only COUNTER JR1 reports: full text downloads at the journal title level. Though this makes the ERM a natural fit for BGSU’s e-journal usage reporting, it could not, when the library was developing this workflow, be used to archive, compile and report searches, sessions and cost-per-use for the library’s databases. Rather than implement yet another tool to house this information, the library decided to keep using spreadsheets to compile database use while lobbying Innovative Interfaces to add this functionality to the ERM. In Summer 2012, the company announced that, in its next release, ERM would be able to work with additional COUNTER report types, including Database 1, 2, and 3 reports. Therefore, the library will be able to more fully automate the process of database reporting beginning in 2013.

Cost-per-download reports for e-journals by title and package have been available from the Innovative ERM for several years, and are precisely the statistics that have always been most labor-intensive to compile on a fiscal-year basis. Throughout 2011, librarians and staff at BGSU worked together to lay the groundwork to move e-journal reporting into the ERM. It was necessary to:

- create and/or update resource records reflective of all active e-journal packages. Each record required a “resource ID” variable field that matched the journal package name in Serials Solutions.
- check and update the payment dates for the invoices recorded in the order records for each subscription package.
- relate the order records for each subscription package to the resource records for those titles (thus creating the connection for cost data).
- complete the coverage load process, a load of all active serials titles by package (the file with this information is provided by Serials Solutions). This process included determining the match point for each title, updating a config file that contains the title and resource ID of each package, and determining the workflow for uploading the report monthly.
- create and test the SUSHI configuration for all compliant vendors. A great disappointment was learning that Innovative Interfaces has only integrated nine of the over 30 SUSHI-compliant vendors into its ERM, and only five of those vendors matched SU’s local subscriptions.
- upload remaining vendors’ JR1 reports into the ERM.
- collect and upload JR1 reports for all vendors back to 2007, where available.

As a result, the library has been able to largely automate the process of generating fiscal-year e-journal statistics with cost-per-use by title for nearly 20,000 journal titles from over 30 different publisher packages without implementing any additional software.

Though using the Innovative ERM for this work is a huge improvement over the library's former, manual methods of generating cost-per-use for e-journals, it is far from perfect. Comparing statistics brought into the ERM via the SUSHI protocol to JR1 reports downloaded from those vendors' websites revealed that the SUSHI counts were inaccurate; therefore, the library has returned to manually uploading those vendors' JR1 files to the ERM. Few vendors' XML reports upload successfully to the ERM, so it is necessary to save all JR1 reports as CSV files and use an online tool developed and maintained by the University of Nebraska-Lincoln to convert them to XML before uploading them to the ERM ([n.d.]). Because of irregularities in the application of the COUNTER standard, many CSV files require some editing before they can be converted. It is complicated to generate accurate cost-per-download information for journal packages where access is available from several sources; for BGSU, this is most problematic for journals that are purchased through OhioLINK and available both in the OhioLINK Electronic Journal Center (a statewide e-journal archive) and from the vendor website (where they typically experience higher use). Finally, the Innovative ERM only produces reports by journal package and, though detailed title-level usage for the journals in each package is available, it must be manually merged to produce a report of all usage by journal title, regardless of package. Reporting from the Innovative ERM is further limited by not offering additional evaluation metrics such as impact factors, etc., something many assessment products do offer (Paynter, 2009).

7. Conclusion

In 2002, Peters wrote, "Two things are needful to make usage statistics of e-resources truly useful: broadly adopted standards and continual practice" (p. 41). Since the development of COUNTER and SUSHI, more and more vendors have adopted these standards, and more assessment products and ERMs are being designed to work with them. However, there are lingering inconsistencies in the application of COUNTER and SUSHI, as well as many vendors who have not adopted them. As more e-products are developed and marketed to libraries, more collections are shifted to electronic format, and more ways to discover and connect to e-content evolve, more challenges related to collecting and interpreting e-resources usage statistics are also surfacing. Libraries need a solid theoretical framework to approach the project of collecting and reporting e-resources usage statistics that can absorb and react to the constantly changing statistical landscape.

This is where continual practice is key. Like Peters, Shim and McClure have recommended libraries make an organizational commitment to working with use data – having adequate staff with adequate knowledge and adequate organizational resources to satisfactorily complete this work. Most libraries, they wrote in 2002, are unprepared to deal with e-resource usage statistics and "simply do not have a culture of evaluation that supports the assessment effort needed to use vendor-based statistics successfully" (p. 512).

Many authors, including Fleming-May and Grogg, have called for libraries to move beyond processing usage counts to using such counts to evaluate the impact of collections (2010a). Luther was already eager to move on from quantitative to qualitative evaluation of e-resources usage when she wrote her report for CLIR in 2000 (2001, p. 132). According to Peters, libraries in 2002 were still studying indicators of use, not use itself (p. 44). With libraries still struggling in 2012 to capture these indicators of use, the shift to studying use itself has still not occurred as hoped for, even ten years later. Development of sound workflows that incorporate existing standards and can be adapted to third-party tools will help the library community create a culture of continual practice around e-resources usage statistics that will allow us to finally move towards studying use itself.

References

- Anderson, I. (2009). *Whither ERMI? the once and future DLF electronic resource management initiative becomes NISO ERM data standards review*. Retrieved May 21, 2012, from http://www.niso.org/apps/group_public/download.php?document_id=2942
- Baker, G., & Read, E. J. (2008). Vendor-supplied usage data for electronic resources: A survey of academic libraries. *Learned Publishing*, 21(1), 48-57. doi:10.1087/095315108X247276
- Blecic, D. D., Fiscella, J. B., & Wiberley Jr., S. E. (2007). Measurement of use of electronic resources: Advances in use statistics and innovations in resource functionality. *College & Research Libraries*, 68(1), 26-44. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=23820121&site=ehost-live&scope=site>
- Blecic, D. D., Fiscella, J. B., & Wiberley, S. (2001). The measurement of use of web-based information resources: An early look at vendor-supplied data. *College & Research Libraries*, 62(5), 434-453. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=eft&AN=502878909&site=ehost-live&scope=site>
- Bordeaux, A., Kraemer, A. B., & Sullenger, P. (2005). Making the most of your usage statistics. *Serials Librarian*, 48(3), 295-299. doi:10.1300/J123v48n03_11
- Collins, M. (2008). Electronic resource management systems (ERMS) review. *Serials Review*, 34(4), 267-299. doi:10.1016/j.serrev.2008.09.011

Collins, M., & Grogg, J. E. (2011). Building a better ERMS. *Library Journal*, 136(4), 22-28. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=59162934&login.asp&site=ehost-live&scope=site>

Coombs, K. A. (2005). Lessons learned from analyzing library database usage data. *Library Hi Tech*, 23(4), 598-609. doi:10.1108/07378830510636373

Duy, J., & Vaughan, L. (2003). Usage data for electronic resources: A comparison between locally collected and vendor-provided statistics. *Journal of Academic Librarianship*, 29(1), 16-22. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=9406179&site=ehost-live&scope=site>

Emery, J., & Stone, G. (2011). *TERMS: What is TERMS?* Retrieved May 21, 2012, from <http://6terms.tumblr.com/post/9997650918/what-is-terms>

ERMI 2 Steering Group. (2006). *DLF electronic resource management initiative, Phase II*. Retrieved May 21, 2012, from <http://old.diglib.org/standards/dlf-erm05.htm>

Fleming-May, R. A., & Grogg, J. E. (2010a). Assessing use and usage: The imperative (the concept of electronic resource usage and libraries). *Library Technology Reports*, 46(6), 5-10. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=tfh&AN=53540866&site=ehost-live&scope=site>

Fleming-May, R. A., & Grogg, J. E. (2010b). Practitioner responses on the collection and use of usage statistics (the concept of electronic resource usage and libraries). *Library Technology Reports*, 46(6), 28-34. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=tfh&AN=53540865&site=ehost-live&scope=site>

Fons, T. A., & Jewell, T. D. (2007). Envisioning the future of ERM systems. *Serials Librarian*, 52(1), 151-166. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=26112186&site=ehost-live&scope=site>

Fry, A., & Najacht, K. (2011a). *BGSU database statistics wiki*. Retrieved May 21, 2012, from

<http://dbusestats.bgsu.wikispaces.net/>

Fry, A., & Najacht, K. (2011b). *BGSU wiki for e-journal usage statistics*. Retrieved May 21, 2012, from

<http://ejusestats.bgsu.wikispaces.net/>

Hulbert, L., Roach, D., & Julian, G. (2011). Integrating usage statistics into collection development decisions. *Serials Librarian*, 60(1-4), 158-163. doi:10.1080/0361526X.2011.556027

Hyland, P., & Wright, L. (1996). Using statistics for database management in an academic library.

Australian Academic & Research Libraries, 27(3), 169-181. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=ISTA3305460&site=ehost-live&scope=site>

Jewell, T., Aipperspach, J., Anderson, I., England, D., Kasprowski, R., McQuillan, B., McGeary, T., Riggio,

A. (2012). *Making good on the promise of ERM: A standards and best practices discussion paper*.

Baltimore, MD: NISO. Retrieved from

http://www.niso.org/apps/group_public/download.php/7946/Making_Good_on_the_Promise_of_ERM.pdf

Liu, W., & Cox, F. M. (2002). Tracking the use of e-journals: A technique collaboratively developed by the cataloging department and the office of libraries technology at the University of Louisville. *OCLC Systems & Services*, 18(1), 32-39. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=ISTA3702145&site=ehost-live&scope=site>

Luther, J. (2001). White paper on electronic journal usage statistics. *Serials Librarian*, 41(2), 119-148.

Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=6214975&site=ehost-live&scope=site>

Matthews, T. (2009). Where does it come from? Generating and collecting online usage data: An update.

Serials, 22(1), 49-52. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=37295808&site=ehost-live&scope=site>

McQuillan, B., Fattig, K. M., Kemp, R., Stamison, C., & England, D. (2010). Electronic resource

management system integration strategies: Opportunity, challenge or promise? *Serials Librarian*, 58(1-4), 106-116. doi:10.1080/03615261003622734

Moen, W. E., Oguz, F., & McClure, C. R. (2004). The challenges of nonstandardized vendor usage data in a statewide metasearch environment: The Library of Texas experience. *Library Quarterly*, 74(4), 403-422. Retrieved from

<http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=16628928&site=ehost-live&scope=site>

National Information Standards Organization. (2008). *Introduction to release 3 of the COUNTER code of practice for journals and databases*. Retrieved May 21, 2012, from http://www.projectcounter.org/r3/r3_intro.pdf

Parker, K., Robertson, N. D. M., Anderson, I., Chandler, A., Farb, S., Jewell, T. & Riggio, A. (2004). *Report of the DLF electronic resource management initiative Appendix B: Electronic resource management workflow flowchart*. Retrieved May 21, 2012, from <http://www.library.cornell.edu/elicencestudy/dlfdeliverables/DLF-ERMI-AppendixB.pdf>

Paynter, R. A. (2009). Commercial library decision support systems: An analysis based on collection managers' needs. *Collection Management*, 34(1), 31-47. doi:10.1080/01462670802548585

Peters, T. A. (2002). What's the use?: The value of e-resource usage statistics. *New Library World*, 103(1172), 39-47. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=ISTA3701832&site=ehost-live&scope=site>

Portland State University. (2011). *ECR workflow*. Retrieved May 21, 2012, from http://dl.dropbox.com/u/17916186/ECR%20Workflow.%20Usage%20Statistics_2011oct.pdf

Shepherd, P. T. (2004). COUNTER: Towards reliable vendor usage statistics. *VINE: The Journal of Information & Knowledge Management Systems*, 34(4), 184-189. doi:10.1108/103055720410570975

Shepherd, P. T. (2006). COUNTER: Usage statistics for performance measurement. *Performance Measurement and Metrics*, 7(3), 142-152. doi:10.1108/14678040610713101

Shim, W., & McClure, C. R. (2002). Improving database vendors' usage statistics reporting through collaboration between libraries and vendors. *College & Research Libraries*, 63(6), 499. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=lih&AN=8744391&site=ehost-live&scope=site>

Tucker, C. (2009). Benchmarking usage statistics in collection management decisions for serials. *Journal of Electronic Resources Librarianship*, 21(1), 48-61. doi:10.1080/19411260902858581

University of Nebraska-Lincoln. ([n.d.]). [*Stats converter for ERM*]. Retrieved May 23, 2012, from <http://statsconverter4erm.unl.edu/>

Walker, M. (2009). E-resource statistics: What to do when you have no money. *Journal of Electronic Resources Librarianship*, 21(3), 237-250. doi:10.1080/19411260903466426

Way, D. (2010). The impact of web-scale discovery on the use of a library collection. *Serials Review*, 36(4), 214-220. doi:10.1016/j.serrev.2010.07.002

