Connecting Electronic Resource Management Systems and Usage Statistics

Oliver Pesch[1]

One of the many, and indeed one of the biggest, challenges libraries face with electronic resources is the fact that these collections are neither housed in, nor accessed from within, the physical library. Unlike print collections, where the library can control access and thus closely monitor use, the electronic content is usually in the control of an outside hosting site—in fact many hosting sites. Libraries who want to monitor usage are required to first fetch statistical data from all of these sites (assuming such reports are provided). Complicating things further is the way e-resources are purchased—a journal may be found in an e-journal package as well as one or more aggregated databases; therefore getting an accurate sense of usage can be difficult as statistics from a number of hosting sites must be consolidated. Fetching, reformatting, loading and consolidating usage to provide meaningful results is both error-prone and extremely time-consuming. This talk will look at how ERM systems provide the structure for organizing electronic resources, and how standards like COUNTER and SUSHI help improve quality and greatly reduce the time and effort of providing effective usage reports.

This paper is about usage statistics and the role an Electronic Resource Management (ERM) system can play in making usage statistics meaningful and effective inputs for collection development and budget management. The need librarians have for statistics on the usage of online collections continues to grow as more and more library information is made available on line. Being able to obtain and process meaningful usage data was not always the norm—many may argue that it is still not the norm. Before addressing ERM systems and how they make use of usage data, some background discussion of the current state of relevant standards and their evolution may be relevant—how we got to where we are and where we need to go.

Background

When online journals and full-text databases began to gain prevalence in the 1990s, libraries started asking vendors to provide usage reports as part of the service to allow monitoring of their investment. Some vendors and publishers resisted for fear that libraries might cancel the online format "if they knew how little it was being used." Other vendors, like EBSCO Publishing, Elsevier and others, responded with a full complement of reporting options. Also during the

[1] Chief Strategist, EBSCO Information Services.

1990s, many library consortia were branching out to provide online information—or were even being created solely for the purpose of purchasing online resources. For them, the need for usage data to justify purchases was a requirement. And it was the library consortia that were, if not the first, likely the most vocal in demanding standardization in usage data. In response to such demand, the International Coalition of Library Consortia (ICOLC) released guidelines for usage statistics on web resources, revised version of which were subsequently published in December 2001 and September 2006 (ICOLC 1998 and 2006).

The significance of the ICOLC guidelines cannot be underestimated. After their release, the full-text database vendors that were not already providing usage data scrambled to do so. The ICOLC guidelines essentially declared that if a vendor wanted to do business with a consortium, they had to provide usage statistics.

Even with the influence of ICOLC, the uptake of publishers in providing usage data for their online journals was somewhat slow. To complicate matters, those statistics that were being provided were sometimes inconsistent and often not comparable between publishers or vendors. As an example, a simple misinterpretation of how a web server records PDF activities could result in gross over-counting of full text downloads—most PDFs display a page at a time and every time a page is requested by the Adobe Acrobat Reader the server records the action; therefore, a single download of a ten-page document could be recorded as one download by one publisher or as ten full-text downloads by another. To address this problem, a group made of librarians, publishers and aggregators was formed in 2002. The work of this committee resulted in the COUNTER (Counting Online Usage of Networked Electronic Resources) Code of Practices, first released in 2003 (COUNTER 2005).

The COUNTER Code of Practice strives for usage statistics that are "consistent, credible and comparable". The code clarifies terminology, the format of the reports, how usage data need to be processed, expectations for delivery, and it provides for an audit. Release 2 of the COUNTER Code of Practice was introduced in 2006 and in the same year a similar code of practice for e-books and reference materials went into effect.

COUNTER's significant achievements include the acceptance among the publisher community that providing usage data was an expectation. It also gave a sense of optimism to the individuals whose job it was to create consolidated

usage reports. People responsible for producing reports for library management and funding bodies had a very challenging job. Before COUNTER, the effort was almost unimaginable as it involved gathering reports from different vendors, then interpreting the information provided so that appropriate numbers could be extracted and put into a summary report. With COUNTER, the consistent formatting made the job much easier and also inspired some librarians to consider how they could automate the process.

All this is happening around the same time as the Electronic Resource Management Initiative (ERMI), out of which today's Electronic Resource Management (ERM) systems were born. ERM systems (as they have become to be known) provide a structure that effectively describes online collections in terms of vendors, packages, host platforms, and resources. This structure made the ERM system a natural consolidation point for vendor usage data. All a librarian would have to do is to retrieve COUNTER reports from each vendor, then load it into the ERM system. The system could then use its knowledge base to produce the needed reports. It all should be very easy.

It should have been easy, but it turns out it was not. Attempts to automate the consolidation of usage data via COUNTER reports uncovered further problems. Even though COUNTER was very explicit in the formatting for spreadsheets, there was enough room for interpretation that different reports from different vendors could not be read by the same import program. And even if the format was consistent, the amount of effort needed to *retrieve* the reports was excessive – this is evidenced by many libraries having lengthy and detailed documentation for fetching and loading COUNTER reports.

Figure 1: One library's procedures for fetching usage data ***(courtesy Ted Fons, Innovative Interfaces).***

In the summer of 2005, a committee was formed to deal with the problem of file consistency and the effort to gather the usage reports. This initiative came to be known as the Standardized Usage Statistics Harvesting Initiative, or SUSHI (NISO 2007). The vision was to create a web service model to allow usage consolidation applications to automatically harvest usage reports from vendors and to require the reports be formatted as XML. By the end of the summer of 2006, SUSHI was a draft NISO standard with many successful implementations. As an example of the time-savings, the Virtual Library of Virginia (VIVA) consortium reported that, for one vendor, they spent ten hours a month obtaining the usage data—with SUSHI this time drops to mere minutes.

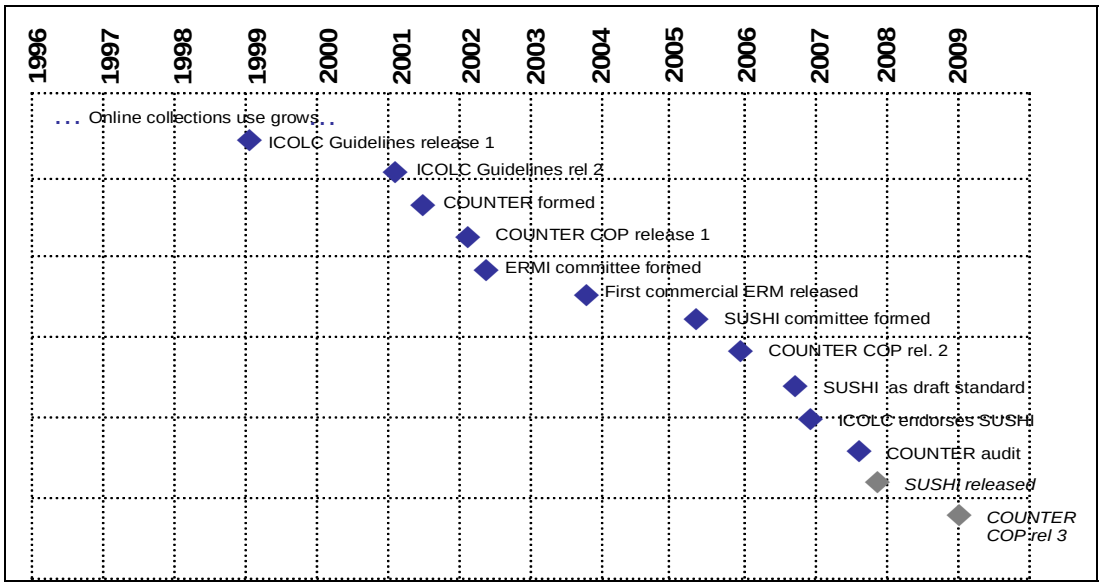

Figure 2: Time line for usage statistics standards and initiatives.

This short history demonstrates an evolutionary process, as each need was addressed, another was uncovered and the community came together with another solution. With SUSHI becoming a full NISO standard in 2007, and COUNTER coming out with Release 3 of their Code of Practice in 2008, the future of ERM systems and other usage consolidation solutions looks promising.

ERM Systems and Their Data Structures

As their name suggests, Electronic Resource Management systems are all about managing e-resources. Those involved with ERMI were frustrated with the confusion caused by managing e-resources – the need to track licenses, trials, URLs, coverage data, rights for and restrictions of use and more – and thus were seeking to provide a single place for a library to store all relevant information about e-resources.

At the heart of the ERM is the knowledge base that organizes the resources in the collection by mapping them to entities such as publishers, vendors (the organization selling the resource), packages or databases (the resource may be acquired as part of a package), interfaces (the platform where the item is accessed), as well as the purchase contract. With this organization of resources, it is possible to answer questions like:

- Which packages and databases are available from a given vendor?
- Which resources are included with a database or package?
- In which databases and packages will a given resource be found?

- On what platforms (interfaces) can a database or package be accessed?
- What databases or packages can be accessed on a given platform?
- What resources can be accessed on a given platform?
- What resources were included in a given deal?
    - From what vendor(s)?
    - On what platform?

Figures 3 and 4 provide a simplistic view of some of the entity relationships. For those interested in a more detailed view of ERM entity relationships, Nathan Robertson contributed a comprehensive diagram that was included as part of the ERMI report (Jewell et al. 2004).

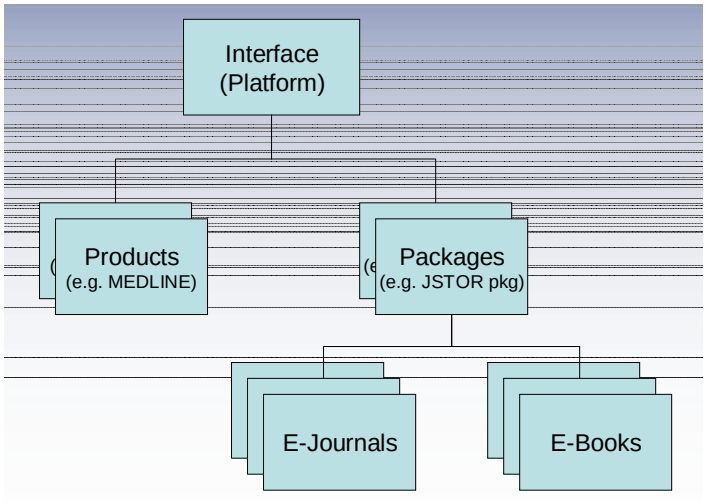

Figure 3: Relationships between interfaces, products and packages.

Figure 3 shows how an interface may offer access to packages of e-journals and e-books. An interface may also provide access to products. Depending on the ERM system, the individual resources that make up a product like MEDLINE may or may not be tracked.

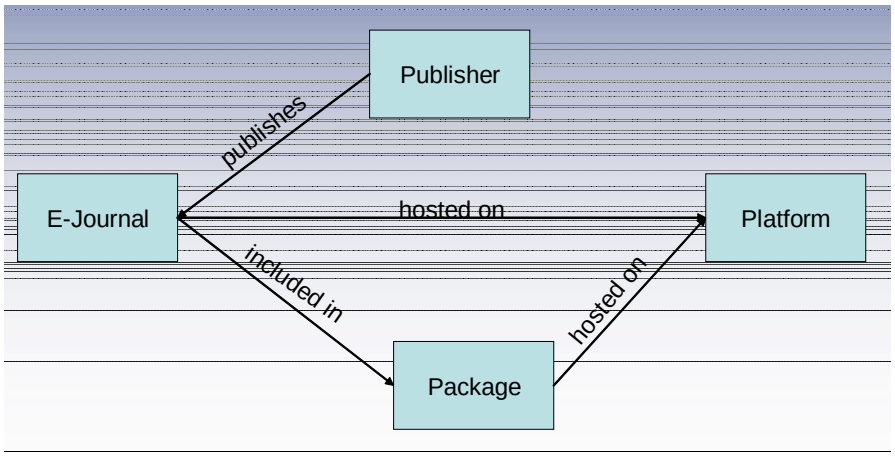

Figure 4: Online journal relationships.

Figure 4 shows how a journal, which is published by a publisher, is hosted on a platform (or interface) and also could be included in a package which, in turn, is hosted on a given platform.

Even though these diagrams are not comprehensive, they demonstrate the need for data in ERM systems to be organized through the use of such relationships in order to be effective.

With the entity relationships in place, the ERM system can associate other information with these entities. Attributes such as:

- Descriptive data (e.g. package names, resource names, etc.)
- Identifiers (ISSNs, ISBNs, package identifiers, etc.)
- URLs (to link to the platform; the database; the actual resource)
- Rights or restrictions for use of the content
- Coverage information
- Cost details
- Usage information
- And much more…

A lot can be written about the data structure of an ERM system and the challenges a library faces in managing their knowledge base; however, this paper is focused on usage statistics so we will limit the discussion to the opportunities and challenges of ERM systems and usage reporting.

ERM Systems and Usage Statistics

Usage data can be a valuable input for managing budgets, collections and the configuration of the systems used to access the collection. Because of the distributed nature of the online collections, the ERM system makes a logical central place to store and consolidate usage data. Once gathered, a number of interesting and useful reports are possible.

- ***Full text usage reports*** provide detailed information about which journals are most used in the collection. The ERM system allows the report to combine usage data from e-journals and aggregated databases to get not only the detailed view of the most used journals, but on which platform was the use.
- ***Cost per use analysis*** focused on e-journals (you probably want to exclude aggregated databases from this report) provides an interesting measure of "value" for a journal. The cost-per-use (CPU) is a simple calculation of subscription cost divided by number of uses. By adding columns for cumulative usage (%) and cumulative budget (as both a value and a percentage) and sorting the report by cost per use (lowest cost first), the report becomes an effective tool to relate the expenditures to usage – for example, you may find that 20 percent of the budget pays for the journals that account for 80 percent of the usage.
- ***Link-out reports*** are another way of getting a sense for how the collection is being used. Unlike full-text usage, which is actual usage as monitored by the vendor, the link-out statistics track the activity of patrons accessing the collection via the A-to-Z list or link resolver integrated with the ERM system. Reports can be pulled to show usage by journal (for collection management) as well as activity by sources and targets of the links or by type of link (configuration management).
- ***Database reports*** usually measure the number of searches conducted for each database. Search counts are a good measure of how effective a database is. For full-text databases, the number of full-text documents accessed from within the database is another good measure of value.
- ***Overlap Analysis*** reports are usually provided as a means of spotting redundancy in a library's collection by highlighting the areas where one database may overlap in coverage with other databases and packages in the collection. The overlap report should consider coverage overlap in addition to title overlap—for example, a given database or package may include a title that is available elsewhere in the collection; however, it may contribute unique years of content. By combining usage and cost data

with the overlap it is easier to make informed collection management decisions.

Here are some examples of reports from Innovative's ERM system. In figure 5, you can see how cost-per-use (CPU) has been included.

	A	B	D	E	F	G	H	I	J	K
1	Number of Successful Full-Text Article Requests by Month and Journal									
2	Year-to-Date 2006 CPU: $4.94									
3	Year-to-Date 2006 Cost: $973,200.58									
4	Year-to-Date 2006 Usage: 196682									
5										
6	Title	ISSN1	2006 CPU	2006 Use	2006-Mar	2006-Apr	2006-May	2006-Jun	2006-Jul	Totals
1222	Surface And Coatings Technology	0257-8972	$23.55	276	56	114	23	34	49	276
1223	Surface Science	0039-6028	$202.93	64	11	12	15	16	10	64
1224	Surface Science Reports	0167-5729	$12.66	10	0	6	1	0	3	10
1225	Surgery	0039-6060	$5.28	24	7	3	6	8	0	24

Figure 5: Full text requests by month and journal ***(courtesy Ted Fons, Innovative Interfaces).***

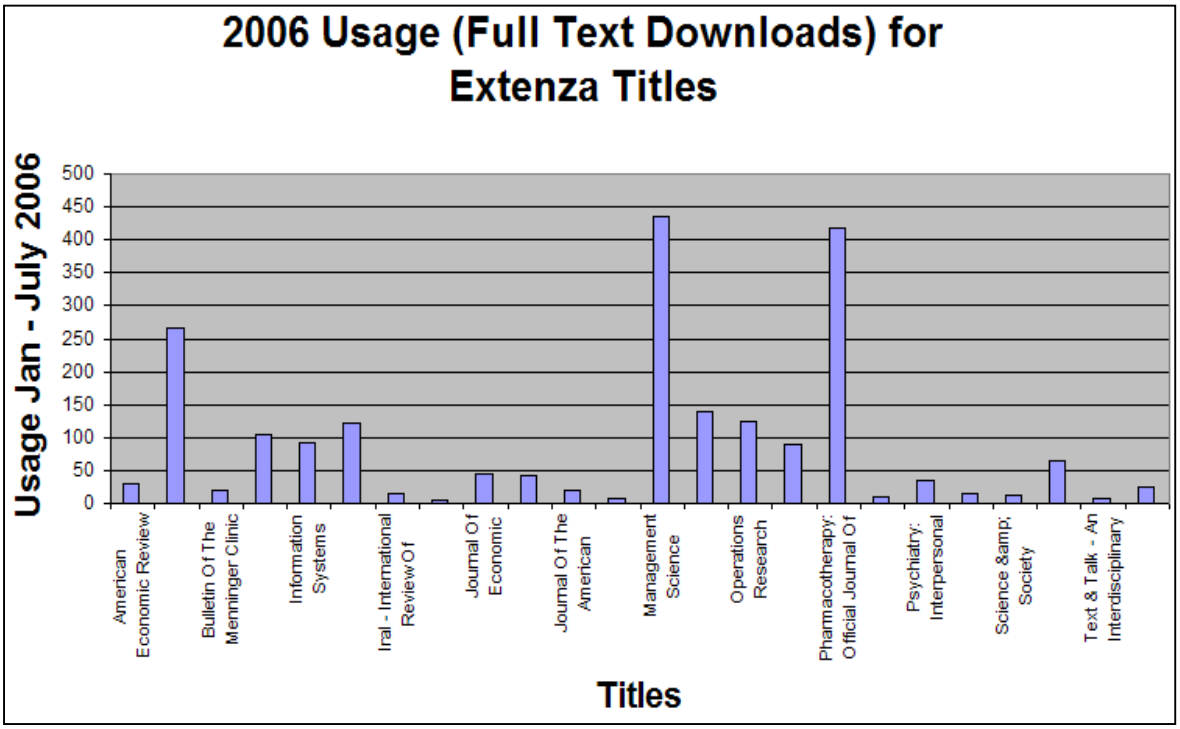

Figure 6: Comparative chart of full text requests by journal ***(courtesy Ted Fons, Innovative Interfaces).***

These are just a few of the possibilities for ERM usage reporting. To make these reports possible, the usage data from the various vendor sites must first be loaded into the ERM system—and it is here that many challenges lie.

ERM Systems and Loading Usage Data

As mentioned in the introduction, the usage consolidation functionality of most ERM systems is centered on COUNTER reports. Journal Report 1 provides details on full text usage at the journal level for a given platform. Database Report 1 provides search statistics by database and platform.

	A	B	C	D	E	F	G	H	I	J	K
1	Journal Report 1 (R2)	Number of Successful Full-Text Article Requests by Month and Journal									
2	Sample University										
3	Date run:										
4	2005-04-05										
5		Publisher	Platform	Print ISSN	Online ISSN	Jan-2005	Feb-2005	Mar-2005	YTD Total	YTD HTML	YTD PDF
6	Total for all journals		EBSCOhost			3942	10006	11093	25041	15776	9265
7	Harvard Business Review	Harvard Business S	EBSCOhost	0017-8012		1117	707	702	2526	1591	935
8	Scientific American	Scientific Americar	EBSCOhost	0036-8733		92	230	221	543	342	201
9	Reading Teacher	International Readi	EBSCOhost	0034-0561		11	190	179	380	239	141
10	Economist	Economist Newspa	EBSCOhost	0013-0613		33	95	155	283	178	105

Figure 7: COUNTER Journal Report 1.

If you look at the COUNTER Journal Report 1 shown in figure 7, you can see month by month statistics shown for each journal. The "Platform" column identifies where this usage occurred, and basic descriptive information is provided to help identify the journal.

Comparing Journal Report 1 in figure 6 and the entity diagram in figure 3, you can see how the usage data "fits" in the ERM system. The usage data in the report will normally be associated with the journal and platform in the ERM system—that is, if a journal, such as *Abacus*, is available on EBSCOhost and on Blackwell Synergy, usage data will be stored for both *Abacus* on EBSCOhost and *Abacus* on Blackwell-Synergy. The key in loading usage data in the ERM is to take the usage information from the report and associate it with the right resource record in the knowledge base. This involves:

- *Matching the platform:* The platform name in the COUNTER report must map to a platform in the knowledge base. Currently there are no standard identifiers; therefore, the ERM system must "know" what text strings to expect in the COUNTER reports.
- *Matching the journals:* The loading process must be able to match a journal on the report to the right record in knowledge base. The report will hopefully provide ISSNs and publisher information in addition to the title of the journal to help in the matching.

On the surface, the loading of usage data should simply be a one-to-one mapping of rows in the usage report to records in the ERM system. However, there are some challenges to be considered:

- *Platform names not found.* The ERM knowledge base must include the platform name or some other mechanism to relate a report to a vendor or host site, otherwise the report can be loaded and the usage will not be included.
- *Variant or missing ISSNs.* As anyone who has worked with serials for any amount of time has already discovered, the ISSN is not really a unique identifier. A common problem in loading the usage reports is that a journal in the report cannot be matched to that journal in the knowledge bases – perhaps because the wrong ISSN is being used, or because the knowledge base represents a different version of the journal (e.g., UK edition versus the North American edition).
- *Duplicate entries on the report.* Some vendors may have the same title appearing more than once in their system – sometimes this is because of how back files are loaded and sometimes this may be because of publisher changes. When duplicates happen, the ERM system should store the cumulative value of the usage.
- *Titles dropped from the knowledge base.* Another common problem that can occur is when titles are removed from a package or database. The ERM knowledge base may drop the title; however, it is still likely that a usage report will continue to report usage for the months where the title was part of the collection. The ERM should allow for this so usage is not lost.

Without a doubt there are other challenges in loading of the data, including the formatting of the spreadsheets. The sophistication of an ERM system comes not only in its ability to provide a well-organized central repository for all information about the library's electronic collection, but also in its ability to gracefully handle exceptions and anomalies in the data to so the library has a complete picture.

Conclusions

In this paper we have shown both the possibilities and the challenges that come with using the ERM system as a usage consolidation tool. Many of these challenges will be mitigated by the adoption of existing standards like COUNTER for consistency in reporting and SUSHI, which automates the harvesting of reports. The introduction of the "linking ISSN" (Reynolds 2006) and similar work for the ISBN plus the creation of services from OCLC like xISBN (one can assume an xISSN service is likely to be coming soon) will help

make resource mapping more accurate. Additional work in the areas of standard identifiers for packages, databases and platforms will also help in managing knowledge bases.

In closing, ERM systems have been created to act as the central place to store much information about the electronic resources in the library collection. Because of the structure and nature of the knowledge base, it make sense to extend the ERM system to capture and consolidate the usage data; thereby usage information can be combined with other data stored in the knowledge base to provide reporting opportunities that are nearly impossible to achieve with any other approach.

References

COUNTER (Counting Online Usage of Networked Electronic Resources). COUNTER Code of Practice for Journals and Databases. Release 2 (April 2005). <http://www.projectcounter.org/code_practice.html> (accessed January 15, 2008).

ICOLC (International Coalition Of Library Consortia). *Statistical Measures of Usage of Web-Based Indexed, Abstracted, and Full Text Resources* (November 1998). http://www.clir.org/pubs/reports/pub94/appendixb.html (accessed January 15, 2008).

_____. *Revised Guidelines for Statistical Measures of Usage of Web-based Information Resources* (October 2006). < http://www.library.yale.edu/consortia/webstats06.htm (accessed January 15, 2008).

Jewell, T.D., et al. *Electronic Resource Management: Report of the DLF Electronic Resource Management Initiative.* Washington, D.C.: Digital Library Federation, 2004. http://www.diglib.org/pubs/dlf102 (accessed January 15, 2008).

NISO (National Information Standards Organization). *Standardized Usage Statistics Harvesting Initiative (SUSHI)* (2007). http://www.niso.org/committees/SUSHI/SUSHI_comm.html (accessed January 15, 2008).

Reynolds, R. "ISSN Revision Draft at Ballot." *NISO Newsline* (January 2006): http://www.niso.org/news/newsline/NISONewsline-Jan2006.htm#Story3 (accessed January 15, 2008).

WorkdCat. XISBN (Web service). http://www.worldcat.org/affiliate/webservices/xisbn/app.jsp (accessed January 15, 2008).