Assessing E-Collections When Every E-Resource Has its Reader, Every Reader has His/Her E-Resource, and E-Resources are Ever Growing

Anjana Bhatt

Florida Gulf Coast University, abhatt@fgcu.edu

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Part of the Library and Information Science Commons

Recommended Citation

Bhatt, Anjana (2006) "Assessing E-Collections When Every E-Resource Has its Reader, Every Reader has His/Her E-Resource, and E-Resources are Ever Growing," Against the Grain: Vol. 18: Iss. 5, Article 7.

DOI: https://doi.org/10.7771/2380-176X.5007

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
Assessing E-Collections When Every E-Resource Has its Reader, Every Reader has His/Her E-Resource, and E-Resources are Ever Growing

by Anjana Bhatt (Electronic Resources Librarian, Florida Gulf Coast University) <abhatt@fgcu.edu>

Electronic resources, or e-resources, whether born digital or converted, are enticing students and librarians alike and have quickly become the format of choice. Libraries subscribe to e-collections ranging from databases to online journals to eBooks, and deal with multiple e-books/e-journal publishers. It is not unusual for e-resources librarians to work with hundreds of databases that contain thousands of e-journals and to manage the nerve-wracking assignment of keeping track of e-collections usage.

Assessment Challenges

E-resources librarians struggle with issues such as content overlap, the need to support multiple formats (some e-journals are only available to those who also subscribe to the print version), constant changes in holdings, lack of perpetual access to electronic information, managing e-journal portals, registering with numerous publishers for online information, negotiating individual licensing agreements, keeping abreast of the growth in e-resource management tools, and last but not least, assessing e-collections on a continuous basis.

At Florida Gulf Coast University (FGCU) Library, we currently provide remote access to 97,000 online journals, of which 46000 are unique titles. We obtain eBooks through NetLibrary, Gale Virtual Reference Books, Oxford Online, and Greenwood Press, and we are in the process of adding My I-Library. In addition, we subscribe to a networked CD-ROM financial software package called Research Insight (formerly known as CompustatPC Plus) for the business faculty and the students in the accounting and financial programs.

As an e-resources librarian, I have faced some thought-provoking challenges in assessing our e-collections. The philosophically nagging question is “what are we measuring, what is being used, who is using the content, how do the databases are being used and can the usage data be a good measure of effectiveness of the databases?” (Luther) Does the e-resource that we are assessing cover the same content/subject matter/chronology as its print counterpart? Unless there is similarity in content, how do we decide the value of one over another? How does one select a particular e-collection when there is considerable overlap among aggregators? Are there good reasons to purchase an e-product even if it is only 50 percent unique?

For consortially-held databases, usage statistics alone cannot drive individual purchasing decisions. What if some libraries are required to participate so the consortium gets a better deal? Even though the database may not be used heavily at your library, consortium membership requires individuals to choose the good of the whole. Canceling a database that is used less often at your institution will jeopardize the price for all the other institutions. Libraries face a strong need to demonstrate the so-called Return-on-Investment (ROI) by maximizing the number of resources while containing costs. How do we determine the ROI for consortial purchases?

E-journal packages pose another set of difficult issues. These packages are created based on instruction activities. How can libraries make with publishers. Are all the journals in the package valuable, or are some just filler? Is it worth spending the money for the valuable journals and just ignore the “thrown ins?” Usage data for these packages is problematic as well. In the absence of statistical data broken down by title, how effective are collection-level usage statistics? If title-level data is available, what does the contract permit? Can you drop journals based on usage data? Unless the journal aggregator allows libraries to add/cancel journals in the package every year, usage data is not very useful. Furthermore, statistics do not usually distinguish among printing, downloading, or emailing, but if they did, should we give greater importance to any one type of use, or user? For example, it is tempting to cancel an expensive research title that is used by just a few faculty members in favor of high-use popular journals (Luther), but that expensive title may be critically important to your faculty. Quantitative data out of context is not useful for answering these questions.

We can all identify with similar difficult, non-usage driven decisions, such as when we are required to spend a large amount on one database because it is critical to one program. Print collections in academic libraries are not generally needed based on usage; should an e-resource be treated differently? Every e-resource like every book might have a user. If usage is low, it may mean that we need to enhance user awareness and increase bibliographic instruction activities. Low usage may also be an indication of a poor user interface, information which should be relayed to the vendor in hopes of improving user usability.

E-resources are ever growing and there is a constant change in the content as publishers lose or add licenses. Usage statistics are affected due to mergers. The e-product market is changing more quickly than our standards, and there is a possibility that Iocolc guidelines and COUNTER Code of Practice may be obsolete for databases that are being developed with new technologies. (Shim & McClure)

Role of Usage Statistics Data

In a study conducted by Shim and McClure (2000), it was found that, in general, libraries are not prepared to commit the necessary resources, staff time, training and effort into data gathering and evaluation. Some studies indicate that librarians are reluctant to employ vendor-supplied usage statistics in assessing and making collection decisions, but other studies conducted at a later date confirm that such vendor-supplied statistics are considered valuable and are, in fact, very much in demand by librarians (McDowell & Gorman). Interestingly, this study reveals that there is less consensus among vendors as to the key types of usage statistics than there is among librarians. Even though librarians and vendors share a significant number of concerns regarding usage statistics, a study (Luther) revealed that vendors use usage statistics as a marketing tool and prefer to collect data only on those factors that can be used to promote the purchase of their products. McDowell and Gorman’s study found that there is no significant relationship between the type of usage statistics provided by the vendors and those desired by academic librarians.

As we all know, funding has increasingly been tied to performance measures (Goldberg). As libraries divert more money every year towards resources, the need to find (in itself difficult) and to interpret this data to assess e-collections in meaningful ways has become more pressing. Where e-resources are concerned, there is a perception that we must use usage statistics to secure and allocate funding, to select and purchase e-collections, to provide ROI justifications to university officials, to understand user behavior patterns and, in some cases, to persuade individual departments to pay part of the subscription cost.

Problems of Using Usage Data for Assessing e-Collections

To state that usage statistics data, at this juncture, is a problematic tool for assessing and comparing e-collections is not an exaggeration (Shim & McClure). Literature review reveals that there are a number of problems associated with operationalizing usage data for assessment purposes. There is no common data set used across all types of e-collections (e-journals/e-databases/eBooks/CD-ROM-based networked products). Data is delivered in inconsistent formats (html, excel or email) and may count few but not all pertinent factors, such as number of total sessions, total hits but no details on individual journal/database usage, full-text access, most visited pages, search sessions, access to articles by journal title/topic/categories, page views, and download requests by formats etc. Vendors use different terminology and may determine on page 26 <http://www.against-the-grain.com>
fine “usage” in such various ways as accessing the homepage of the e-resource, downloading a full text article, or simply accessing table of contents/abstract/or full text (McDowell & Gorman). Keeping in line with protecting user’s privacy they also avoid collecting data on user behavior. Most vendors provide information on number of sessions or searches but in reality a user might have logged in repeatedly to do the same search for a variety of reasons (system failure, interrupted work session, too many simultaneous users), so this information is of dubious utility.

Time period of usage statistics is another concern. Some vendors (e.g., StatUSA) do not keep statistics for more than six months on their Website and do not even bother to provide the previous data on demand. Most vendors provide the ability to choose a time period for data retrieval, but in some cases it is not possible to choose a specific time period or to pull up data for more than one year in a single report in order to analyze the changing usage patterns. One has to be content with whatever time periods are available. There are different standards for the time limit and the number of clicks that vendors count may result in the counting of one usage as multiple usages. It takes considerable time to check individual e-resource’s Website and collect this information for presenting it in a suitable format for ensuring a proper analysis. Merging this data together for a realistic analysis in the absence of qualified staff is another problem that many libraries face. Most libraries prefer to access the raw data available on the vendor’s Website but hesitate to compile it for comparison purposes as they are not sure about what would best suit the institutional needs. Even though standardization is a critical factor, vendors have to realize individual needs of libraries and must provide abilities to customize the data (McDowell & Gorman).

How realistic is it, then, to assess e-collections with the help of—unfortunately ever so inadequate and difficult to measure—usage statistics data? During my tenure of just one and a half years as an e-resources librarian, I have come across three incidents of vendors reporting inaccuracy of their usage data. When libraries have no control in collecting data and can not identify consistent under, or, over counts (Luther, Shim & McClure) how wise is it to trust the validity of the usage statistics for making informed decisions? Until usage statistics produce meaningful data that can be interpreted in a sense capable ways, libraries cannot depend solely on this information to make collection assessment decisions.

The broader question, however, is this: how much do usage statistics really matter when every reader has his/her resources and every e-resource has its user? At best, the usage statistics for any given e-resource tell nothing more than how many articles were downloaded or how many search sessions were initiated. Did the user find what he/she was looking for? Do hit counts reflect on the quality of the articles? And for growing universities, is it realistic to stop subscribing to costly, possibly lesser-used databases that might be required for program accreditation purposes, for important faculty research, or which are critically important to one or two programs? For example, at FGCU, we pay $20,000 for Research Insight software because this e-resource has its users

Continued on page 28

Against the Grain / November 2006

<http://www.against-the-grain.com>
and there are some users who will not be effective if we stopped subscribing. We, however, have collected usage data through the Fall and Spring semesters in 1998-99 and persuaded the College of Business to allot $5000 towards the subscription.

Usage Statistics, ScholarlyStats and COUNTER Code of Practice

Librarians, vendors and publishers all have realized the value of usage data for their own specific purposes and there have been a number of initiatives in this direction. The most promising is the Counting Online Usage of Networked Electronic Resources (COUNTER) Code of Practice. It is a developing international standard that is jointly owned by publishers, intermediaries, libraries, consortia and industry organizations. Designed for all of its participants and with an objective to develop a standard code of practice for measuring usage of online information products, COUNTER is credible and compatible. Comparison of usage data is the key element for assessing e-collections and COUNTER fulfills this need by providing a consistent framework for recording and exchange of online usage data. Along with the standardized definitions of data elements, it also includes standards for data elements to be measured, content and format of usage reports, requirements for data processing and auditing, and guidelines to avoid duplicate counting.

It is encouraging to note that in addition to a revision of first code for journals/databases (including detailed specifications for consortium-level reporting), Release 1 of the COUNTER Code of Practice for books and reference works was issued in March 2006. This first version brings the possibility of ensuring comprehensive industry standards for the recording and reporting of usage statistics for eBooks. It includes a set of six basic usage reports that cover full-text requests for a whole title, as well as for sections within a title such as chapter, page, and paragraph. Searches, sessions and turn-aways are covered as well. As usual it also includes the glossary of terms relevant to books and reference works. Interestingly, the format of the report is almost consistent with the existing codes for journals and databases, as the specifications for report delivery, data processing, audit and compliance are identical. The only exception is that the content of the usage reports has been changed and the set of definitions of terms has been expanded. Participants can access usage data from a password-controlled Website. Users can request email alerts on updated reports, which are available in various formats (CSV, Excel and XML) and provided on a monthly basis. New reports are available within four weeks of the end of the reporting period.

Most importantly, the data from the previous year is also available, which is critical for comparison purposes. At this time, however, it is not possible to pull data for more than one year into a single report to analyze usage patterns. Another drawback is that currently very few vendors are COUNTER compliant (Luther, Shim & McClure) and libraries that may want to use COUNTER reports are forced to deal with two sets of data. Some vendors, in spite of being COUNTER compliant, do not provide the ability to extract reports on their Websites and require you to sign up for receiving these reports on a monthly basis. Unfortunately, COUNTER Code of Practice does not prescribe a delivery method as well.

To alleviate some of the problems associated with collecting usage statistics, libraries can now look forward to using a new tool: ScholarlyStats. Launched in January 2006, and marketed by MPS Technologies, ScholarlyStats is designed to manage and evaluate usage statistics from multiple sources (Alliances & Deals). Currently, ScholarlyStats covers 39 vendors and promises to deliver monthly, ready-to-analyze consolidated journal and database usage reports in standardized COUNTER-compliant format. Usually librarians need to go to a particular Website and extract (pull) statistics; potentially, the best part of subscribing to ScholarlyStats is that statistics are pushed to the client in a neat, unified package. Recently three major US consortia (Michigan Library Consortium (MLC), PALINET and the Carolina Consortium) have purchased ScholarlyStats. ScholarlyStats reports can also be used by publishers to analyze usage patterns, improve user interface/navigation problems, and rectify such access issues as logout time and access cut-off.

Pricing for ScholarlyStats, however, is a matter of concern. Their current cost structure is based on the number of platforms and vendors from which they collect reports for a library. Presently, the pricing for 9,24, and 25-39 vendors is $3,995, $9,720 and $14,270 respectively. Faced with constant funding crunches, how many libraries can afford to spend this amount for collecting usage statistics? It would be interesting to conduct a study that could calculate the staff time in terms of money and compare this cost against the cost of subscribing to ScholarlyStats. All said and done, though, ScholarlyStats and COUNTER standards are a glimmer of hope for the near future.

Practical Solutions

As stated earlier, it is currently impossible to compare usage statistics data across a number of e-collections; however, I have found that this data can be used to a certain extent for assessing e-collections. Currently, I manually collect annual statistics for about 55 databases, out of which only ProQuest and EBSCO databases are COUNTER compliant. For our consortium databases, I depend on the usage data that is broken down by institution and is gathered by our consortium administrator. Since "Sessions run" and "Searches run" are the most commonly available statistics, I track our databases on these two factors and our technical services head has plans to run this data against "cost per usage" price to determine ROI and also to identify "low-usage", high "cost per usage" databases. When collection librarians evaluate new databases or make renewal decisions, I use the overlap analysis tool available through the client center of the Serials Solutions product to provide unique content information and title lists for informed decision-making. Usage data can perhaps be an indicator of databases for poor search interfaces as well. In the future we would like to collect data on the highest number of hits for journals that provides only abstracts and no full-text articles so we can subscribe to these titles online.

We do recognize that e-resources are ever growing and have tremendous effect on customer services and collection process in academic libraries. We are making conscious efforts to replace paper with e-content and assess e-collections while keeping in mind that every user has his/her e-resource and every e-resource may have its user. At this time, we fervently hope that more and more e-vendors would opt to be COUNTER compliant and provide standardized usage statistics data for realistic assessment of e-resources.

References


<http://www. against the grain.com>
Digital Library of Earth Systems Education: Collections Assessment

by Barbara DeFelice (Director, Digital Resources Program, Dartmouth College Library) <Barbara.DeFelice@dartmouth.edu>

and Constance Rinaldo (Librarian of the Ernst Mayr Library, Museum of Comparative Zoology, Harvard University) <crinaldo@oeb.harvard.edu>

BACKGROUND

The Digital Library for Earth System Education (DLESE) is a multi-faceted digital library that includes a searchable collection of metadata records for educational resources. Our work in DLESE focused on the Educational Resources Collection component of this digital library. When we refer to the DLESE collection in this paper, we mean this Educational Resources Collection. DLESE was open to public use in 2000 and continues today; our project examined data from spring of 2002 through December 2004. Aspects of the metadata, search system and collections workflows and policies have changed over time, so we will use current terminology in this paper, although some of the processes we describe have been modified since this project ended.

The collection was developed for teachers and learners interested in all aspects of the earth system. Records describe objects at different levels of granularity, such as a whole course, a syllabus, an image, a collection of images, a data set, a tutorial or an applet that illustrates a scientific concept (Kastens et al., 2005). When the DLESE search page is opened (http://www.dlese.org), a user can choose to search on a word or phrase or to use non-subject qualifiers (such as Grade Level and Resource Type). A user can also browse by subject or collection, and can limit subject term searches by Grade Level, Resource Type, Educational Standard and Collection.

DLESE describes its collection along the dimensions of Subject (originally called Topic), Grade Level (originally called Learning Context), and Resource Type (originally called Learning Resource Type), among other metadata elements (DeFelice, Kastens, Rinaldo, & Weatherley, 2006; Rinaldo & DeFelice, 2005). Metadata assigned to a resource must include one or more grade levels, resource types, and controlled vocabulary topics, terms, which can be used to browse or to limit a free text subject search. Therefore, these attributes were used in our collection assessment project.

Collection assessment is the process of systematically comparing the scope and balance of a library’s actual collection with the scope and balance of materials desired by the library users (Hall, 1985; Nisonger, 1992; Richards & Eakin, 1997). Materials desired by the users, staff, funding entities, and/or overseers of the library are described in the collection policy. Materials desired are also expressed in the actions of users, by the types of materials requested by the community, from direct user feedback, and through searches for materials. In the case of DLESE, information about users’ desires comes from analysis of requests to the Search and Browse functions of the DLESE Discovery System.

Collection assessment becomes more critical when a collection is being built for an unknown audience. DLESE users are not required to register to search the Educational Resources component of DLESE, so there are no records of exactly who uses DLESE, how much, and how satisfied they are with the collection. Without a central collection development operation, there is no guarantee that the collection will be balanced in terms of depth and breadth. We did not assess individual resources in a collection, nor the effectiveness of individual items in improving science education. Our assessment was designed to try to identify gaps in the collection so that the intended audience could be served. Since the intended audience was “anyone interested in learning more about the Earth,” the collection needed to satisfy any age level, people with different educational backgrounds, teachers and students in formal and informal learning environments, and people from a wide range of geographic areas (Manduca & Mogk, 2000). The collections assessment was intended to help guide DLESE collection developers by providing general directions for proactively building a useful collection for this broad audience.

DLESE was planned from the start to have a volunteer, contributor-based collection development program to support a broad scope statement and a diverse potential user group. The collection needed to be balanced, broad and of high quality for this very diverse, and to a large degree unknown, user group. Therefore, some means of monitoring the collection in order to guide the collection building efforts was needed. Otherwise, the collection might reflect only the interests of the early adopterscontributors, and fail to meet the needs of many new users. We wanted to find something useful in the DLESE collection so they would be motivated to use it.

When a library user finds no resource(s) that match a query, the user experiences a negative interaction with the library. Many factors contribute to this interaction, including user confusion about the search system, metadata or other kinds of indexing that does not match the user’s inquiry, or lack of content that meets the needs expressed in the search. We wanted to see if we could isolate the areas where a lack of content was the key issue.

METHODS

Data Acquisition

Working with the staff at the DLESE Program Center (DPC), we identified the kinds of data and the formats for this data which we thought would be most useful at the beginning of the project. There are many different kinds of data to use, so it took a considerable amount of time to isolate the data that would give us the insights we wanted and to decide what filters to apply to the data. Working closely with the technical staff at the DLESE Project Center, we continually checked the data for consistency and quality over time.

Digital libraries change constantly, and since the project started, some changes in topic terms took place, and other useful metadata was added, such as Educational Standards. However, we continued to focus on the key parameters for assessment that we picked in the beginning of the project, Subject, Resource Type and Grade Level. Search and browse logs show what users are looking for (user behavior), and searches that have no results (null searches) help identify potential collection gaps and perhaps provide information about where the search engine and user behavior do not match (Gagnon & Makuch, 2003). We mapped free text searches to the Subject metadata provided by DLESE. We reviewed and characterized zero-result (null) searches to determine if we could identify gaps in the collection to aid collection builders in developing balanced collections. We analyzed search logs to identify reasons for zero result (null) searches. The search engine was revised over time and these revisions changed responses for some of the observed user queries during the course of our study. As more resources were added search responses also changed.

continued on page 32

Accessing E-Collections ... from page 28


<http://www.against-the-grain.com>