



Making sense of e-book usage data

Title	Making sense of e-book usage data
Author(s)	Cox, John
Publication Date	2007
Publisher	The Acquisitions Librarian

Making Sense of E-book Usage Data

John Cox

ABSTRACT. This article provides an overview of the types of statistical data available for e-book usage and includes examples from specific vendors, along with coverage of standards such as COUNTER (Counting Online Usage of Networked Electronic Resources). It describes ways in which usage data can be employed for purposes such as monitoring of e-book uptake and collection development. Limitations with regard to vendor usage statistics are highlighted, and the importance of supplementing them with surveys and other studies of user behavior is emphasized, with examples.

KEYWORDS. Electronic books, usage statistics, user studies

INTRODUCTION

The range of e-books available for purchase or subscription has expanded significantly in recent times and, though uptake continues to lag e-journals, they now form an important element in libraries' e-resource collections. The very fact that uptake is variable according to discipline and other circumstances creates a strong imperative for libraries to study usage data carefully. There are many unanswered questions regarding the impact of e-books and, as for other e-resources, librarians need to understand patterns of usage in order to monitor value for money and to guide collection development decisions. Ideally they would enjoy easy access to a comprehensive array of usage data consistently collected by different vendors. Unfortunately, this is not currently the case, and

John Cox is Deputy Librarian, National University of Ireland, Galway, Ireland (E-mail: john.cox@nuigalway.ie).

The Acquisitions Librarian, Vol. 19(3–4), 2007
Available online at <http://www.haworthpress.com>
© 2007 by The Haworth Press. All rights reserved.
doi: 10.1080/08963570802026278

librarians need to be prepared for some frustration and the investment of an amount of extra effort in establishing the effectiveness of their e-book collections.

Despite the fact that e-books have been around for some time, they remain less established than e-journals, particularly in terms of subscription models that vary widely but also with regard to usage analysis. For e-journals, there is consensus that the full-text article provides a consistent unit of measurement (Shepherd, 2006). E-book products are more diverse, and possible metrics include whole book titles or individual chapters, sections, and entries. In addition to the diversity of offerings from vendors, there are different data collection practices, while delivery of usage data is also complicated by digital rights management issues. Efforts at standardization are very welcome but face many challenges. This article examines what is currently possible in the area of e-book usage analysis and describes some encouraging initiatives. It argues throughout, however, the need to inform the deployment of vendor-supplied usage statistics with a keen awareness of possible limitations and the value of supplementary approaches such as surveys in understanding the actual usage and impact of e-books.

E-BOOK USAGE STATISTICS

What types of usage data can we expect from e-book vendors? (A list of e-book vendors is provided in the Appendix). As already mentioned, there are many variations between vendors, but Table 1 outlines the more common metrics, and Figures 1 and 2 show sample reports. These metrics are not universally reported by vendors. Early English Books Online (EEBO) and Oxford Reference Online are examples of services that do not show usage by title, whereas session duration is not included in NetLibrary or Eighteenth Century Collections Online (ECCO) reports. Additional reports may include usage by subject (NetLibrary, ebrary), unsubscribed books browsed (Safari, Ebook Library [EBL]), details of pages accessed within a book (EBL) or copied/pasted from a book (informa-world [Taylor & Francis]), activity by IP subnet (ebrary), and number of unique users (EBL). Commonly available facilities include limiting reports by time period and output of data to spreadsheet software for further manipulation. However, not all services provide a self-service administration module, and it may be necessary to request reports from vendors.

TABLE 1. Common E-book Usage Metrics

Metric	Explanation	Notes
Sections viewed, printed or downloaded	The number of full-text content units viewed, printed or downloaded by users.	Content units may be whole books, chapters, vendor-defined sections or, for reference works, individual entries.
Usage by title	The number of times a title was accessed.	May be the total use of constituent parts, e.g., chapters or sections, which are counted separately.
Number of sessions	The number of user sessions (i.e., period from opening to closure of access).	Sessions can also be counted as ended at timeout following a specified period of inactivity.
Duration of sessions	The total length of time for all sessions.	Average session length may be given separately or can be calculated.
Number of turnaways	Frequency with which users are denied access.	Usually because the maximum number of permitted concurrent sessions is in use.
Searches	The number of searches conducted by users.	Supplementary data may include details of search terms used or the number of searches returning zero hits.

Applications for E-book Usage Statistics

Library staff employ e-book usage data for a number of purposes, primarily in the contexts of monitoring levels of use and making collection development decisions.

Monitoring Use

Usage statistics enable varying levels of e-book uptake measurement. Raw numbers can at least show patterns of use for any given product although, as discussed later, comparison between products is more difficult. Thus, Dillon and Langston provide examples of the analysis of title and subject data for early NetLibrary subscriptions at the University of

FIGURE 1. *NetLibrary* Usage by Title (<http://www.oclc.org/netlibrary/>)

Accesses	Title	Subject	LCC	Dewey	Authors	Publisher	ISBN	Year	eContent ID	Content Type
163	Monitoring for Health Hazards At Work	Medicine	R41229.A84 1999eb	615.9 /02	Ashton, Indira.; Gill, F. S.	Blackwell Science,	9780632050413	2000	51411	eBook
148	Politics in the Republic of Ireland 4Th Ed.	Political Science	JN1415.P65 2005eb	320.9417	Coakley, John.; Gallagher, Michael	Routledge in association with PSAI Press	9780203411810	2005	116438	eBook
113	Environmental Hazards : Assessing Risk and Reducing Disaster	Social Sciences: General	GB5014.S6 2004eb	363.34	Smith, Keith.	Routledge	9780203354315	2004	94678	eBook
109	International Law 5Th Ed.	Law	KZ3275.S53 2003eb	341	Shaw, Malcolm N.	Cambridge University Press	9780521824736	2003	125137	eBook
63	Public Administration and Public Policy in Ireland : Theory and Methods	Political Science	JN1435.P83 2003eb	320 /.S /09417	Adshead, Maura.; Millar, Michelle	Routledge	9780203411551	2003	96997	eBook
54	Political Ideologies : An Introduction	Political Science	JA71.H49 2003eb	320.5	Heywood, Andrew.	Palgrave Macmillan	9780333961773	2003	101548	eBook
51	The Evolution of EU Law	Law	KJ6947.E99 1999eb	341.242 /2	Craig, P. P.	Oxford University Press,	9780198264811	1999	18319	eBook

Texas at Austin (Dillon, 2001a) and California State University (Langston, 2003), respectively. Safley reports data for a wider range of e-book services and genres at the University of Texas at Dallas, with significant increases between 2004 and 2005 (Safley, 2006). Cox noted the role of data in identifying peaks and troughs in uptake of the Safari service at National University of Ireland, Galway (Cox, 2004b), while Christianson and Aucoin examined trends in monthly use of NetLibrary titles at Louisiana State University for a 13-month period (Christianson and Aucoin, 2005).

The latter study is one of a number employing usage data to compare the use of e-books with their printed equivalents. Littman and Connaway studied activity for 7,880 titles in print and via NetLibrary at the Duke University Libraries over the period from February 2001 to August 2002 (Littman and Connaway, 2004). They noted a key deficit of earlier studies that compared the number of times an e-book was accessed with the number

FIGURE 2. *EEBO* Usage Summary Report (<http://eebo.chadwyck.com/>) (Image published with permission of ProQuest. Further reproduction is prohibited without permission)

Date	Web Sessions	Searches	Citation Views	Illustration ToC Views	Document/Page Image Views	PDF Downloads	Full Text (ASCII) Views	Searches Returning No Hits	Turnaways
May 2007	75	212	37	5	164	21	61	93	0
Apr 2007	277	1091	146	15	747	156	295	567	0
Mar 2007	285	880	106	59	1848	252	179	429	0
Feb 2007	133	361	62	40	896	202	119	146	0
Jan 2007	398	333	84	10	588	163	124	111	0

of circulations for its printed equivalent. This tended to operate in favor of e-books that might attract multiple accesses over a short period of time, while a single circulation tended to take up a relatively long period. Furthermore, in-library use of printed books did not feature. Their study therefore compared *whether* the books had been accessed or borrowed rather than the frequency of transactions. They found that the e-books involved received 11% more use than their printed counterparts in the study period and had the potential to deliver good value. There were also some interesting patterns of overlap and format preference, with 39% of titles being used in both formats, 34% as e-books only and 27% used only in print. Christianson and Aucoin focused on 2,852 titles available in both formats at Louisiana State University and found some differences, with higher use of printed books but greater concentration of e-book usage in fewer titles (2005).

The analysis by Safley (2006) at the University of Texas at Dallas, already mentioned, found that print circulations had decreased from 67,465 in 2004 to 50,993 in the first 11 months of 2005, a period during which usage of a variety of e-book services had increased significantly. This indicates a gain for e-book uptake at the expense of printed books. Bailey (2006) reported the same trend at Auburn University at Montgomery Library where printed book usage declined by almost a third from 2000 to 2004, and use of NetLibrary titles increased considerably, reaching almost 23% of the level of printed book usage.

Uptake according to discipline was a common focus of all these studies and others besides. Littman and Connaway (2004) found that e-books were used more than their printed equivalents in education, medicine, psychology and computing, whereas Safley observed strong uptake overall, notably in computer science and engineering but also in history (Safley, 2006). Relative to its representation in the collection of NetLibrary titles at California State University, Langston (2003) reported strong use in computer science and technology and engineering. Among the subjects more fully represented, economics and business, medicine, sociology, and psychology received the greatest number of accesses. Business and computing, along with literature and medicine, feature prominently in Bailey's figures for e-book usage at Auburn University itself but also in his table of comparative use at a number of universities based on raw numbers of accesses (2006). The similarity in subject profile among a number of institutions suggests that e-books are best suited to certain subjects.

Collection Development and Management

Data on uptake by discipline and the impact of e-books on the use of printed titles inform collection development decisions (e.g., influencing the balance between print and online book purchase in certain subjects or identifying titles where online access may satisfy demand better than multiple printed copies). A survey of librarians by ebrary Inc. [Ed. Note: the survey appears elsewhere in this volume] asked respondents to indicate decisions influenced by e-book usage statistics (ebrary Inc., 2007). Collection development issues accounted for three of the four choices offered and ranked highest, with renewal and budget allocation both at 59% and title acquisition not far behind at 53%.

Value for money is a key factor, and usage data can provide some insight in this respect. Wilkins describes the deployment of usage statistics to assist in renewal decisions at the University of Derby and the use of spreadsheets to record how much each title has cost over a number of years (Wilkins, 2007). Taylor-Roe and Spencer (2005) highlight the reporting of investment in, and usage of, e-textbooks acquired for a newly established distance learning course at Newcastle University. Safari offers a book-swapping facility, and Cox notes the benefit of dropping two low-use titles in favor of two new titles on Java, one of which established itself as the most used book in the subscription at National University of Ireland, Galway (Cox, 2004b). Where provided (e.g., EBL, NetLibrary, Safari [Fig. 3]), details of attempted access to, or previews of, unsubscribed titles can help to identify priorities for acquisition. Data regarding turnaways can also help to monitor the adequacy of concurrent user license provision, as highlighted by Cox and Dillon (Cox, 2004b; Dillon, 2001a).

E-book usage data offer the opportunity to check the effectiveness of content selection procedures. Thus, Safley compared usage of a

FIGURE 3. *Safari* Previews Report (<http://www.safaribooksonline.com/>)

Hit Type	Total Hits (Views)	Title
Preview	155.00	Knowledge Management Toolkit, The
Preview	82.00	RESTful Web Services
Preview	70.00	A First Look at SQL Server 2005 for Developers
Preview	51.00	UML for Mere Mortals®
Preview	46.00	How Personal & Internet Security Work
Preview	44.00	Sams Teach Yourself Java™ 2 in 21 Days, Third Edition
Preview	43.00	Python Programming on Win32
Preview	41.00	Refactoring: Improving the Design of Existing Code

librarian-selected e-book collection (NetLibrary) with that of a vendor-supplied collection (ebrary) and discovered almost identical usage by subject, raising some questions about the impact of selection activity (2006). Williams and Best found that favorable *Choice* reviews had little impact on the likelihood on e-book circulation which was actually higher for non-*Choice* titles, although the sample at Auburn University at Montgomery was very small (2006). The effectiveness of cataloging is another possible focus, exemplified by Dillon's reporting of the impact of cataloging of e-book titles on patterns of uptake at the University of Texas at Austin (Dillon, 2001b). Encouragingly, the addition of catalog records appeared to have a positive immediate effect and resulted in a significant increase in use with changes in uptake by subject, too. Less positively, an earlier analysis by Dillon found that many high-use titles had previously been acquired in print but were now missing (Dillon, 2001a).

Other Uses

While collection development issues dominated the ebrary survey mentioned earlier, a third of the respondents also made use of e-book usage statistics to plan training and promotion (ebrary Inc., 2007). Data on usage can be used to influence focusing of information literacy initiatives for a given e-book service or to plan training according to discipline. Allied to this, it may be possible, depending on the level of reporting offered, to get some insight into how e-books are being used. Data on most frequently accessed titles or sections may show a high concentration on a small number of readings, suggesting a focus on recommended texts; conversely, a wider distribution of titles might indicate supplementary reading by students. The number of pages viewed and the volume of printing may offer some indication of levels of onscreen reading or of a tendency to use an e-book service simply as a delivery mechanism for hardcopy access. Equally, session duration may suggest different levels of use from casual browsing to intensive research or simply quick reference. Safley reported session length and activity data for ebrary, Safari, EEBO, *Oxford English Dictionary* and Oxford Reference Online (Safley, 2006), while Cox adduced some possible learner behaviors from *Safari* statistics (Cox, 2004a).

In reality, raw usage data can offer only limited assistance with understanding e-book user behavior and tend to support speculation rather than authoritative comment. As outlined later in this article, there is wide recognition of the need to supplement usage statistics with deeper studies of actual use. It is perhaps telling that 14% of the respondents to the

ebRARY survey indicated that usage statistics had no influence on decisions, possibly reflecting some of the deficits in data provision discussed next.

Difficulties with E-book Usage Statistics

Some of the difficulties with e-book usage data are common to e-journals, and others are specific to e-books, or at least far more pronounced than for e-journals. This is certainly true of one of the major issues, the lack of consistency between vendors in the data they deliver. Safley is by no means alone in noting that “each publisher provides different variables to approximate the intensity of the use. Comparing statistics between companies is a problem because of the lack of standards” (Safley, 2006).

Reference has already been made to variations in data collection practices, and some examples of data provided by one vendor but not another were given. This has immediate implications for attempts to compare use between services or to get an overall picture of e-book uptake as a whole across the range of subscriptions, which is typical for the majority of libraries. For example, it is unlikely that a library could generate an average session duration for all e-book use because this metric is not supplied by all vendors. Some vendors present the entire book as a single downloaded file whereas others count chapters or entries as separately downloaded sections (Shepherd, 2006). Exactly what constitutes a section is another variable. For Oxford Reference Online “full-content units requested” correspond to the number of discrete entries viewed in its constituent reference works, but five print pages viewed online is the measure for Oxford Scholarship Online. Safari sections typically correspond to three pages of the printed book, though there is no differentiation by section in NetLibrary which simply counts “accesses” to a title.

At the heart of these inconsistencies is the range of different access models offered by e-book vendors and, in particular, restrictions on use generated by digital rights management (DRM) policies. DRM is more complex for e-books than e-journals owing to publishers’ concerns about possible loss of print revenue for academic textbooks especially. Rice provides an overview of e-book licensing models (Rice, 2006) and Ferguson (C. Ferguson, 2006, 2007) summarizes practice with regard to user-permitted printing, downloading, and copy-paste activity. Restrictions such as allowing only the printing of one page at a time (e.g., NetLibrary), not supporting downloading (e.g., ACLS Humanities E-books), or disallowing copying and pasting (e.g., informaworld) inevitably impact on user

behavior. Usage reporting, as well as being subject to variation according to the access model used by different vendors, only reflects the extent of actual permitted use rather than full potential activity. DRM significantly affects opportunities for comparing usage across e-book platforms.

Sessions and searches represent further areas of frustration, although the issues here are not exclusive to e-book services. Vendors may set different timeout periods, effectively resulting in different definitions of a session. Blečić, Fiscella, and Wiberley (2007), in a study of e-resources, generally found variations of from five to 30 minutes among six e-resource vendors questioned. Such variations will affect the numbers of sessions and turnaways counted by different vendors. The same study also noted the tendency for federated searching to inflate the number of sessions counted, resulting in a lowering of the search-per-session ratio. This may be a greater issue for databases but is worth bearing in mind for e-books, too, as is another variable among vendors with regard to the counting or non-counting of zero-hit searches, which was also noted in this study. Reporting of zero-hit searches and the terms used can help in understanding user experience but is rare. Frustratingly, Safari no longer offers this facility whereas EEBO reports the number of such searches but does not indicate which search terms were used.

Returning to search sessions, it is worth remembering that one session may represent many users. The notes on EEBO's usage statistics point out that a session begins when a new user starts to use EEBO and ends when the browser they are using is closed. It is possible, for example, that a public-use workstation could support a number of users in the same session. Furthermore, all sessions, consecutive or otherwise, on such a workstation might only be counted as a single access by some vendors. Duy and Vaughan noted ProQuest's practice of only counting visits from unique IP addresses (2003). Their study excluded e-books but noted variations between local and vendor counting of sessions, as did Ferguson and Chan (A. Ferguson and Chan, 2005), whereas Blečić et al. reported significant differences in session and search counts between vendor-specific and COUNTER-compliant statistics for the same e-resources at the same time periods (2007). The work of Project COUNTER is described later, but the clear message is that interpretation of e-book and other e-resource usage statistics needs to be done with caution and a clear understanding of how usage is measured by the vendor. A further consideration, and one that COUNTER has striven to address, is the possibility of double-counting when a title is available through an aggregator service such as MyiLibrary but its publisher also reports usage.

FIGURE 4. Safari Reports List (Partial) (<http://www.safaribooksonline.com/>)

BackOffice

Safari
TECH BOOKS ONLINE

Accounts Users Reports Help

LogOut

Reports > My Reports

Reports

Quick Search: ID [] Search Advanced Search

Count Results

ID ▲	Template Title ▲	Description ▲	Run ▲
10699	User Listing Based on Account	Lists all users in a specific account. Details include user's name, ID, status and activation date...	Run
22151	User Detail Hits Report	This report gives a listing of all sections viewed, by user, for the previous month.	Run
22170	User Activity Summary	This report will provide a session activity summary by user for any specified account covering the...	Run
22222	Bookshelf Swapping Report	This report lists the books that were added and removed from user's bookshelves during the previous...	Run
22403	Bookshelf Title Listing	(Formerly named "Bookshelf Usage Report" This report lists all books on a user's bookshelf...	Run
23308	Number of Sections Read	This report will give the Number of sections read and title for the current month for the entire acc...	Run
864399	All books contained in Account Library	Lists all books assigned to the Account's Library.	Run
878675	Summary Hits Report	Returns Preview and View Hits for an entire Account, for the current month.	Run
892716	Number of Hits Per User	Totals the number of hits by user for a specified period of time.	Run

Collecting and analyzing e-book usage data is hard work. McLuckie observes that some usage statistics for e-books are sent by e-mail, others have to be downloaded from a vendor Web site, and some may need to be requested by the subscriber (McLuckie, 2005). For vendors who provide a Web site for usage data report generation, there is a need to schedule visits at appropriate intervals and to know a range of usernames, passwords, and URLs. Choosing the right report can be a challenge, as in the case of Safari (Fig. 4), but at least the element of choice is welcome. In other instances, there is only a predefined standard report available and subscribers have very little flexibility. EEBO and Oxford Reference Online are examples, and neither service offers data for individual titles within it, greatly limiting possibilities for analysis.

Downloading to third-party software for further manipulation may be necessary, and Wilkins describes the supplementary use of spreadsheets to keep track of annual cost per e-book title, to remove data for earlier but now lapsed titles, and to analyze accesses on a per-title basis (2007). Spreadsheets may assist with the generation of graphical output, which emerged as one of a number of common deficits in a *Charleston Report* informal survey of more than 100 librarians on electronic usage statistics in 2007 (The Charleston Report, 2007). There is a need to assign the

right staffing resources, both in terms of time and skills, to gathering and interpreting usage data. The survey just mentioned found that the electronic resources librarian or serials librarian usually collects usage statistics and that few respondents allow student employees to undertake this function. Kidd (2005) notes the absorption of significant senior library assistant time in collecting e-resource data at Glasgow University, but Taylor-Roe and Spencer (2005) report dialogue with other libraries that do not have a staff member specifically designated to this role.

A further consumer of time is the need for liaison with vendor company personnel to clarify the specification employed for usage data and to seek improvements. Cox (2004b) reports extensive discussion with ProQuest regarding the usage data provided for Safari, while Duy and Vaughan (2003) observe that it can be difficult to locate a vendor company representative who can advise authoritatively on data interpretation. Identifying the right vendor contact may be a particular issue for e-books, as Milloy also noted a need to consult several departments when seeking project bids from e-book publishers (Milloy, 2007).

Librarians need continually to ask questions of the figures presented to them. A single book or maybe a small number of titles could account for much of the usage of an e-book service. Thus, five titles from a subscription of 64 e-books generated half (2,011 of 4,026 hits) of Safari usage at National University of Ireland, Galway between September 1 2003 and February 8 2004 (Cox, 2004a). A significant proportion of titles had no use or very limited uptake, a factor to be considered in assessing the overall success of the subscription. Langston (2003) notes that the apparently strong uptake of e-books in anatomy and physiology at California State University was almost entirely the result of intensive use of a single title that accounted for 121 of 143 uses in that subject between March and December, 2001. He also advises against the over-reliance on reports by subject in NetLibrary, noting the very general headings used and the assignment of a heavily-used computer software manual to the business and economics category rather than to computer science. Another point to make sure of is that accesses to unsubscribed content in a service are not included in its overall hit counts or can at least be easily separated from the total. Remember also that not all e-books are equal. Different patterns of use can be expected between historic works, such as those included in EEBO, reference titles in Oxford Reference Online, and textbooks in NetLibrary, all reported in Safley's study (2006).

E-book usage data leave many questions unanswered. The number of turnaways is of interest but, with the exception of Safari, there is often no

further information regarding the time of day when they occurred or the number of users locked out at the same time. Information about on- or off-campus use would be welcome, but ebrary is an exception in providing details of calling IP address. As noted at the end of the previous section, the biggest frustration lies in trying to employ usage statistics to understand meaningfully the actual user experience or satisfaction with e-books. How many sessions, hits, or downloads represent success from the viewpoint of the user or indeed the subscribing library? The point is neatly summarized in the summary of responses to the e-resources usage statistics survey in *The Charleston Report* already mentioned: "The stats are purely quantitative. They don't tell you whether the users got the information they were looking for, whether it was enough, too much, etc." (The Charleston Report, 2007). It is necessary to use other methods to get a deeper understanding of e-book usage and users.

BEYOND USAGE STATISTICS

Surveys of e-book users represent the most common way of supplementing the numerical picture provided by usage data. They can provide deeper insights into user demographics, levels of satisfaction, and actual nature of use. California State University, as reported by Langston (2003) encountered limitations with usage statistics and also conducted a voluntary online survey when evaluating NetLibrary during 2001. This survey gleaned information on user profile, location of use, routes to e-book discovery, and satisfaction levels. A generally positive view of e-books emerged. Levine-Clark (2007) used a general survey of awareness and usage of e-books at the University of Denver in 2005 to establish usage patterns specific to the arts and humanities, exploring issues such as discovery, frequency of use, reasons for use, e-book collections used, and format preference between print and online. There was an overall preference for print from this survey population, along with a focus on content rather than software features. The Irish university libraries surveyed users, mainly in IT subjects, of the Safari service in 2004 to assess user satisfaction and reasons for use but with a focus on support for learning activities (Cox, 2004b). Respondents reported high satisfaction with coverage, varying knowledge of software functions, and differences in purpose of use between students and academics. The learning support section of the survey revealed consultation of a wider range of books in Safari than in the library, along with definite views that Safari had improved respondents' work and saved time.

A more recent survey of more than 1,800 faculty at the University of London (Rowlands, Nicholas, Jamali, and Huntington, 2007) forms part of the SuperBook project conducted by the Centre for Information Behaviour and the Evaluation of Research (CIBER) team and featuring more than 3,000 e-books contributed by Oxford University Press, Wiley Interscience, and Taylor & Francis. The survey examined patterns of current e-book use according to age group, gender, and academic status along with format preference, purpose of reading, discovery routes, and awareness of library e-book provision. One of the interesting findings in this particular survey was considerably higher uptake of e-books by males. Conflicting messages emerged regarding format preference, however. In one section of the survey users claimed a strong preference for reading from screen rather than paper but later reported a definite preference for hard copy in terms of ease of use. As noted by the authors, this highlights a weakness in surveys in that they deliver data on reported rather than actual behavior.

CIBER has been at the forefront of research into actual user behavior through its use of deep log analysis, initially for e-journals but now e-books, too. For e-books, Connaway and Snyder (2005) offered a precursor to this work by publishing the results of a transaction log analysis of NetLibrary usage. They studied raw transaction logs from NetLibrary on a given date, February 26, in 2002, 2003, and 2004 and were able to obtain information often unavailable from vendor-supplied e-book usage statistics. This included number of unique users, peak times of day, length of time on the site and within an e-book, and the number of unique pages viewed per session. Extraction of the data proved to be a large and time-consuming effort, and it is not surprising that log analysis has been undertaken by research teams rather than by individuals or by library staff.

The deep log analysis technique developed by CIBER was able to go a step farther by correlating usage data and demographic user data in a study of the Blackwell Synergy e-journals service (Nicholas, Huntington, and Watkinson, 2005). This service required user login, enabling the characterization of users according to academic status, geographical location, and institutional affiliation. The deep log analysis employed in this study examined "site penetration," including types of item viewed, referral link used, and subject of journal searched, along with repeat visitor identification and activity.

Advantages claimed for this technique include its inclusion of all users, as opposed to the tendency in surveys to limit to a sample or self-selected population, and the focus on actual behavior rather than self-reported activity, which may be influenced by a range of factors including questionnaire

design. It is limited, however, in terms of explaining behavior but is effective in identifying issues for further examination through surveys or other methods.

Deep log analysis is now being applied to studies of e-book usage, including the SuperBook project already mentioned, where it has delivered a number of insights into the usage of Oxford Scholarship Online at University College London, as reported for the period January–March 2007 (Nicholas et al., 2007). Compared with e-journals, findings in this particular study included more intensive use and greater consultation of older titles. Cataloged titles were found to be much more likely to be used, while examination of use according to sub-networks enabled some analysis by subject and according to on- or off-campus location.

Another application for deep log analysis and other forms of user study will be in the Joint Information Systems Committee (JISC) national e-books observatory project in the United Kingdom (JISC Collections, 2007). This project commenced in mid-2007 and aims to study e-book usage at U.K. higher education institutions in depth, focusing on a collection of 36 titles, including some of the most popular texts in business studies and management, medicine, engineering and media studies, accessible via the MyiLibrary and Books@Ovid platforms. The project will include a deep log analysis study by CIBER as part of the overall effort to understand e-book user needs and behavior and the impact of e-books on teaching and learning practices and on library print circulation data, as well as to inform promotional activity, pricing, and licensing. There will be student and academic focus groups, interviews with librarians, and a final user questionnaire, making for a multifaceted approach to delivering insights into e-book usage and information-seeking behavior by users.

A further technique to note is direct observation of users. Hernon and others report a study of e-book use by undergraduates in nursing, literature, and economics at Simmons College in Boston (Hernon, Hopper, Leach, Saunders, and Zhang, 2007). This study involved direct observation and interviewing of 15 students who were undertaking an assignment in each of the three disciplines. Participants used a think-aloud protocol to describe what they were doing and why; notes were taken to document their actions, and the interviews provided some further insight. The study encompassed a range of e-book services and probed awareness, reasons for use, services used, and activity such as browsing, printing, downloading, and annotation. Improvements in the layout of e-books by publishers in order to facilitate online reading and in the presentation of resource discovery facilities by libraries emerged as areas for further action.

COUNTER AND SUSHI: A BETTER FUTURE?

Two standardization initiatives promise valuable progress in addressing some of the current difficulties with e-book usage statistics. The first of these is Project COUNTER (COUNTER, 2007). COUNTER initially focused its efforts on e-journals. It built on earlier work by the International Coalition of Library Consortia (ICOLC) which originally published its *Guidelines for Statistical Measures of Web-Based Information Resources* in 1998, with further revisions in 2001 and 2006 (International Coalition of Library Consortia, 2006). ICOLC identified five key data elements to be included in usage reports, namely numbers of sessions, searches, menu selections, full-content units accessed, and turnaways. It also defined delivery parameters, including Web-based access within 15 days of the end of the preceding month. Oxford University Press is an example of an ICOLC-compliant e-book publisher. Unfortunately, the ICOLC guidelines were not mandatory, and vendors chose to interpret them in a variety of ways. In 2003, COUNTER published the first release of its *Code of Practice for Journals and Databases*, followed by Release 2 in April 2006. COUNTER is a collaborative venture between publishers, intermediaries, and libraries. This approach has enabled it to secure agreement on definitions for formerly vaguely used terminology and to get buy-in from vendors for the generation of standard user reports. Vendors are keen to advertise COUNTER compliance, and 100 have met the requirements of the code for journals and databases.

The success of COUNTER with e-journals provided an excellent springboard for an expansion of its work to e-books, leading to the publication of the *COUNTER Code of Practice for Books and Reference Works* in March 2006. As described by Shepherd, the same collaborative approach used for e-journals guided this initiative (Shepherd, 2006). The task was less straightforward than for e-journals owing to the more diverse range of publishing and access models for e-books. As a result, the code has had to take account of situations where only a whole book could be downloaded or where content might be presented in smaller units such as chapters or entries such as dictionary definitions (all grouped by COUNTER under the term *section*).

Its first two reports, therefore, provide separately for the number of title and section requests by month and title, with the remaining reports focusing on the number of turnaways and total searches and sections. For each report there are tight definitions of the terminology used and

specific requirements to be met in generating the usage reports, including protocols applicable when aggregators and gateways are involved. An impressive feature is the requirement for third-party auditing of reports to ensure continuing compliance, reinforcing user confidence in these reports. The 38-page code, supplemented by appendices, also specifies formats for report delivery, access via a password-controlled Web site with e-mail alerts when updates occur, monthly reporting at least, and updates within 4 weeks of the end of the reporting period.

COUNTER at least offers the prospect of comparable usage statistics from a range of vendors. It has to be noted that the approach is to achieve a fairly basic standard that most vendors can meet rather than to push for a fuller range of possibilities that might be delivered only by a smaller range of vendors (Borghuis, 2005). A standard level of comparability is the aim, and its achievement will be valued by librarians. Initial progress has been frustratingly slow, however, with only 11 publishers listed on the COUNTER Web site as compliant with the e-book code more than two years after its publication. Walker, executive vice-president at Credo Reference, one of these publishers, advocates compliance, and urges publishers to “just do it” (Walker, 2007). There are reports of increased interest in the code, and it must be hoped that many more vendors will follow this lead.

The second development of note is closely linked with COUNTER. The Standardized Usage Statistics Harvesting Initiative (SUSHI) promises to reduce significantly the labor involved in retrieving and analyzing usage data (National Information Standards Organization [NISO], 2007). This is a NISO initiative that aims to automate the delivery of COUNTER-formatted statistics. SUSHI, still a draft standard at the time of writing, is a retrieval protocol that should save the need for visiting lots of Web sites and manually downloading usage data on a vendor-by-vendor basis. It is a simple object access portal (SOAP) request-response Web services “wrapper” for the XML version of COUNTER reports. It will work best with an electronic resource management (ERM) system acting as the client, thus facilitating the downloading of usage data into a system already populated with bibliographic, financial, and licensing data. This offers the possibility of automatic analysis of subject, cost, and cost-per-use data, with the potential for more detailed and extensive calculations than heretofore (Hendricks, 2007). The prospects are certainly exciting, but these are early days, and there is a need for further development in terms of support for SUSHI among a critical mass of publishers and also of ERM capability and adoption.

CONCLUSION

Making sense of e-book usage data is a complex process and demands persistence and flexibility on the part of librarians. Assembling the data available from vendors involves considerable effort, but the real work begins only with trying to analyze the data. There is a distinct lack of consistency between vendors in the data supplied, a trend further complicated by the wide range of access and licensing models on offer. DRM practices designed to protect print revenues, especially for textbooks, are a significant influence. Some vendors offer a variety of reports, whereas others deliver a single summary report with little scope for customization. Elements of data provided by one vendor are often not supplied by another, making comparison of e-book usage across a number of services difficult or impossible. Even when data appear to be comparable, it is important to establish how each vendor collects information, to ascertain what is and is not counted, and to clarify definitions for terms such as *sessions* and *downloads*. There remain many unanswered questions in terms of demographics, actual user experience, and levels of satisfaction, and it is not surprising that many libraries have used surveys to supplement usage data, while researchers have developed deep log analysis. Keeping up to date with user studies and projects is an important part of understanding the uptake and impact of e-books.

Despite their limitations, e-book usage data are essential for libraries and play a vital role in monitoring uptake and in a range of collection development and management decisions, often involving significant sums of money in the case of subscriptions or of labor as regards cataloging or promotion and training. Standardization and automation of e-book data collection, as promised by COUNTER and SUSHI, respectively, are prizes well worth fighting for, and librarians need to use their influence to accelerate progress in these domains. It is not enough to bemoan current inconsistencies and inadequacies, and there is a vital lobbying role for librarians to play in seeking improvements and supporting industry-wide initiatives. There are definitely opportunities to exert more influence with vendors by bringing usage reporting requirements to the fore in negotiating subscriptions and renewals. As e-books grow in popularity, it is clear that the importance of consistent, comprehensive and accurate usage data will also increase. Though frustrations will endure in the immediate term, improvements are certainly attainable.

REFERENCES

- Bailey, T. P. (2006). Electronic book usage at a Master's Level I university: A longitudinal study. *The Journal of Academic Librarianship*, 32(1), 52–59.
- Blecic, D. D., Fiscella, J. B., & Wiberley, S. E. (2007). Measurement of use in electronic resources: advances in use statistics and innovations in resource functionality. *College & Research Libraries*, 68(1), 26–44.
- Borghuis, M. (2005). How COUNTER continues to help librarians and vendors make sense of usage reports [Electronic version]. *Library Connect*, 7, 8–9. Retrieved from <<http://libraryconnect.elsevier.com/lcp/0701/lcp070112.html>>.
- Christianson, M., & Aucoin, M. (2005). Electronic or print books: Which are used? *Library Collections, Acquisitions, and Technical Services*, 29(1), 71–81.
- Connaway, L. S., & Snyder, C. (2005). Transaction log analyses of electronic book (E-book) usage. *Against the Grain*, 17(1), 85–93.
- COUNTER (2007). COUNTER: Counting Online Usage of NeTworked Electronic Resources. Retrieved October 26, 2007, from <<http://www.projectcounter.org>>.
- Cox, J. (2004a). *E-books in support of a new learner environment*. Paper presented at the IUISC 2004: Irish Universities Information Services Colloquium. Retrieved October 26, 2007, from <<http://www.iuisc.ie/2004/presentations/Thursday/John%20Cox.pdf>>.
- Cox, J. (2004b). E-books: Challenges and opportunities [Electronic version]. *D-Lib Magazine*, 10. Retrieved October 26, 2007, from <<http://www.dlib.org/dlib/october04/cox/10cox.html>>.
- Dillon, D. (2001a). E-books: The University of Texas experience, part 1. *Library Hi Tech*, 19(2), 113–124.
- Dillon, D. (2001b). E-books: The University of Texas experience, part 2. *Library Hi Tech*, 19(4), 350–362.
- 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.
- ebrary Inc. (2007). *ebrary's global eBook survey*. Retrieved October 26, 2007, from <<http://www.ebrary.com/>>.
- Ferguson, A., & Chan, G. R. Y. C. (2005). Usage statistics at Hong Kong University: From fun to fundamental in just a few years [Electronic version]. *Library Connect*, 7, 4–5. Retrieved October 26, 2007, from <<http://libraryconnect.elsevier.com/lcp/0701/lcp070104.html>>.
- Ferguson, C. (2006). Technology left behind—eBook rollout. *Against the Grain*, 18(5), insert.
- Ferguson, C. (2007). Technology left behind: The second eBook rollout [Electronic version]. *Against the Grain*, 19. Retrieved October 26, 2007, from <http://www.against-the-grain.com/_old/ebookrollout/>.
- Hendricks, A. (2007). SUSHI, not just a tasty lunch anymore: The development of the NISO Committee SU's SUSHI standard. *Library Hi Tech*, 25(3), 422–429.

- Hernon, P., Hopper, R., Leach, M. R., Saunders, L. L., & Zhang, J. (2007). E-book use by students: Undergraduates in economics, literature, and nursing. *The Journal of Academic Librarianship*, 33(1), 3–13.
- International Coalition of Library Consortia. (2006). Revised guidelines for statistical measures of usage of web-based information resources, October 4, 2006. Retrieved October 26, 2007, from <<http://www.library.yale.edu/consortia/webstats06.htm>>.
- JISC Collections. (2007). JISC national e-books observatory project. Retrieved May 12, 2008, from <<http://www.jiscebooksproject.org>>.
- Kidd, T. (2005). Usage statistics and how we're using them: The example of Glasgow University [Electronic version]. *Library Connect*, 7, 3. Retrieved October 26, 2007, from <<http://libraryconnect.elsevier.com/lcp/0701/lcp070102.html>>.
- Langston, M. (2003). The California State University E-book Pilot Project: Implications for cooperative collection development. *Library Collections, Acquisitions, and Technical Services*, 27(1), 19–32.
- Levine-Clark, M. (2007). Electronic books and the humanities: A survey at the University of Denver. *Collection Building*, 26(1), 7–14.
- Littman, J., & Connaway, L. S. (2004). A circulation analysis of print books and e-books in an academic research library. *Library Resources and Technical Services*, 48(4), 256–262.
- McLuckie, A. (2005). E-books in an academic library: Implementation at the ETH Library, Zurich. *Electronic Library*, 23(1), 92–102.
- Milloy, C. (2007). E-books: Setting up the national observatory project. *Library and Information Update*, 6(11), 32–33.
- National Information Standards Organization. (2007). NISO Standardized Usage Statistics Harvesting Initiative (SUSHI). Retrieved October 26, 2007, from <http://www.niso.org/committees/SUSHI/SUSHI_comm.html>.
- Nicholas, D., Huntington, P., & Rowlands, I. (2007). E-books: How are users responding? *Library and Information Update*, 6(11), 29–31.
- Nicholas, D., Huntington, P., & Watkinson, A. (2005). Scholarly journal usage: the results of deep log analysis. *Journal of Documentation*, 61(2), 248–280.
- Rice, S. (2006). Own or rent? A survey of eBook licensing models. *Against the Grain*, 18(3), 28–29.
- Rowlands, I., Nicholas, D., Jamali, H. R., & Huntington, P. (2007). What do faculty and students really think about e-books? *Aslib Proceedings*, 59(6), 489–511.
- Safley, E. (2006). Demand for e-books in an academic library. *Journal of Library Administration*, 45(3–4), 445–457.
- Shepherd, P. T. (2006). The COUNTER Code of Practice for books and reference works. *Serials*, 19(1), 23–27.
- Taylor-Roe, J., & Spencer, C. (2005). A librarian's view of usage metrics: Through a glass darkly? *Serials*, 18(2), 124–131.
- The Charleston Report. (2007). TCR electronic surveys: electronic resource usage statistics. *The Charleston Report*, 11(5), 3.
- Walker, J. (2007). COUNTER: Getting the measure of ebooks [Electronic version]. *eLucidate*, 4, 3–7. Retrieved October 26, 2007, from <<http://www.ukeig.org.uk/>>.

Wilkins, V. (2007). Managing e-books at the University of Derby: A case study. *Program: Electronic Library and Information Systems*, 41(3), 238–251.

Williams, K. C., & Best, R. (2006). E-Book usage and the Choice outstanding academic book list: Is there a correlation? *Journal of Academic Librarianship*, 32(5), 474–478.

Appendix. E-book Services Cited

Title	URL
<i>ACLS Humanities E-Books</i>	http://www.humanitiesebook.org
<i>Books@Ovid</i>	http://www.ovid.com/site/products/books_landing.jsp
<i>Early English Books Online</i>	http://eebo.chadwyck.com/
<i>Ebook Library</i>	http://www.ebilib.com/
<i>Ebrary</i>	http://www.ebrary.com/
<i>Eighteenth Century Collections Online</i>	http://gale.cengage.com/EighteenthCentury/
<i>informaworld</i>	http://www.informaworld.com/
<i>MyiLibrary</i>	http://www.myilibrary.com/
<i>NetLibrary</i>	http://www.oclc.org/netlibrary/
<i>Oxford English Dictionary</i>	http://www.oed.com/
<i>Oxford Reference Online</i>	http://www.oxfordreference.com/
<i>Oxford Scholarship Online</i>	http://www.oxfordscholarship.com/
<i>Safari</i>	http://www.safaribooksonline.com/