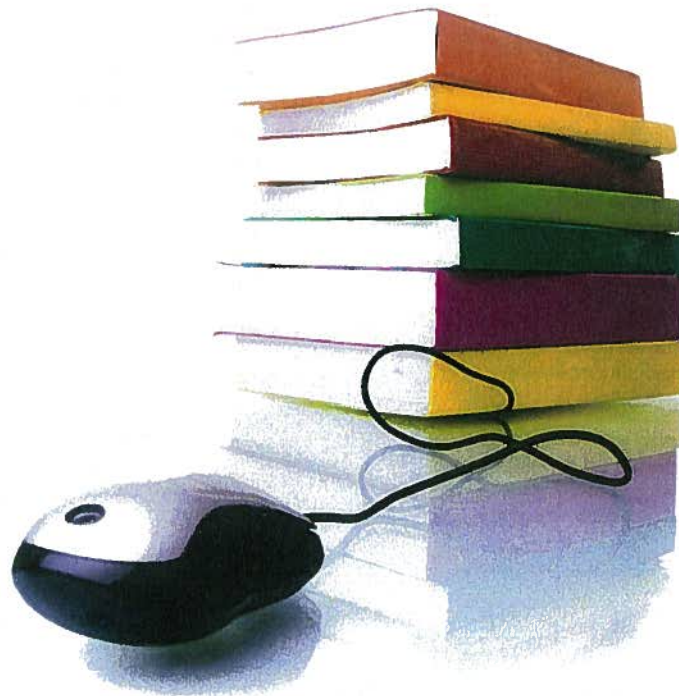


Chandos Information Professional Series

Evaluation of Digital Libraries

An insight into
useful applications
and methods

Edited by Giannis Tsakonas
and Christos Papatheodorou



CP CHANDOS PUBLISHING

Contents

<i>List of figures and tables</i>	<i>vii</i>
<i>About the contributors</i>	<i>xi</i>
<i>List of abbreviations</i>	<i>xix</i>
<i>Preface</i>	<i>xxi</i>
1 Introduction: the framework for digital library evaluation	1
<i>Tefko Saracevic</i>	
PART 1: TO WHOM IT MAY CONCERN	15
2 From usage to user: library metrics and expectations for the evaluation of digital libraries	17
<i>Brinley Franklin, Martha Kyrillidou and Terry Plum</i>	
3 An agency perspective on digital library evaluation	41
<i>Michael Khoo, David McArthur and Lee Zia</i>	
PART 2: WHAT TO PLACE UNDER THE EVALUATION LENS	61
4 What should we take into consideration when we talk about usability?	63
<i>Judy Jeng</i>	
5 Users and digital libraries: an insightful story	75
<i>Emmanouel Garoufallou, Ranla Slati and Richard J. Hartley</i>	
6 Towards an Infrastructure for digital library performance evaluation	93
<i>Maristella Agosti and Nicola Ferro</i>	
7 Employing deep log analysis to evaluate the information-seeking behaviour of users of digital libraries	121
<i>David Nicholas</i>	

From usage to user: library metrics and expectations for the evaluation of digital libraries

Brinley Franklin, Martha Kyrillidou and Terry Plum

Introduction

The refinement of e-metrics by librarians to evaluate the use of library resources and services has rapidly matured into a set of standardised tools and shared understandings about the value of e-metrics for making data-driven, managerial decisions in libraries. These measures grow out of a long history of libraries desiring to assess the usage of their resources, but only a recent history of actually being able to do it. Usage assessment of library collections began with the print materials that were collected, owned, organised and made available by the library. The methods and tools for assessment have multiplied as journals became digital and were no longer owned by the library. The assessment focus has also changed and is now on service as well as collections.

In academic libraries, the frameworks and purposes for assessments of collection usage have substantially changed, driven by the accompanying changes in the content delivery mechanisms, user behaviour, assessment tools and analytics. Not only are the assessment tools and data far better for digital materials than they ever were for print, but the frameworks for collecting, analysing and using those data have changed from the print world to the digital world, and are now consistently used and accepted across academic libraries. The collections are regarded as both a resource and a service. As a resource, assessment focuses on how the collection has been used. As a service, assessment focuses on how the

patron has used the digital resource. This chapter examines the changed assessment framework for digital materials to determine what guides its development and acceptance among librarians. It looks at various projects that attempt to use usage data to learn something about user outcomes.

The scope of this survey is focused on patrons' interactions with networked electronic resources and services. This chapter does not address the current state of the evaluation of digital libraries, which is ably discussed in the other chapters in this book. It does not include web usability or usability analysis, even though the usability of the library's presentation of electronic resources and services affects patron success and therefore usage.

Interestingly, in this evaluative mini-world of networked electronic resources and services, there is a disconnection between what librarians are trying to do with their collection and analysis of evaluative data and the evaluations computer scientists are doing with digital libraries. This chapter surveys current evaluation methods and techniques used for collecting traces of patron interaction with networked electronic resources (i.e. usage data). It also looks at how those data are being used to generate information about user outcomes. The chapter then speculates qualitatively on how this assessment culture in the library world frames librarians' expectations about the assessment evidence techniques and data proposed by information scientists to evaluate or assess digital libraries. The framework of librarians' evaluative expectations is important because most production digital libraries exist within the library and not the computer science context.

Assessment and evaluation in the print library environment

Scarcely a decade into the digital library environment, librarians already know considerably more about digital library use than they did about traditional library use in the print environment. In traditional print-based libraries, librarians counted outputs such as circulating library materials, reference and information questions, and interlibrary loans to and from other libraries. In retrospect, the data collected were not reliable and, most likely, inconsistent due to varying loan periods, local practices regarding how to count informational and directional versus reference questions, and variances in how libraries classified interlibrary loans as opposed to circulation transactions.

Prior to the advent of online catalogues and integrated library systems, even circulation data for specific volumes or classes of materials were difficult to compile and analyse. Books were added to libraries' collections, but detailed information about their subsequent circulation patterns was not easily available until libraries began to automate the library equivalent of electronic inventory control systems in the last quarter of the twentieth century. The data collected by the automation of cataloguing and circulation systems made it easier to count collection size and circulation, and break them out by subject. Journal review projects in the print environment were predominantly in response to budget crises, were undertaken largely to cancel titles that were perceived as less frequently used and/or that seemed overpriced, and were frequently subject to manipulation by either librarians/selectors or the users they consulted.

Before the emergence of digital libraries during the last decade, librarians evaluated collection usage data, when they were: (a) interested in measuring library performance; (b) asked to compile statistics for professional associations or governmental agencies; or (c) confronted with budget cuts and needed to determine how the collection was being used. Librarians typically relied on gross circulation counts and routinely employed unscientific and unreliable sampling plans and simple in-house data-collection methods such as asking users not to reshelve library materials so the library could count them. These usage studies purported to measure library collection use when in fact there was never any tangible proof or consistent interpretation of what a book being removed from the shelf, or even a circulating item, really represented.

Looking back, collection development in the print environment was more of an art than a science. Libraries knew how much they were spending, but were unable to ascertain how their collections were being used or how to use the data they could collect to better inform purchasing decisions.

It is telling that the authors of one of the most commonly cited articles on print collection use in an academic library, 'Use of a university library collection', observed that:

...the gross data available up to this point have been too global in character and too imprecise in nature to serve as an adequate basis for the reformulation of acquisitions policies. It is not particularly helpful for a bibliographer to know that 10 per cent of the titles selected will satisfy 90 per cent of client demand for materials in a given discipline, unless we can determine which 10 per cent. It is

useless to tell the acquisitions librarian that half the monographs ordered will never be used, unless we can specify which 50 per cent to avoid buying. (Galvin and Kent, 1977)

Automated library systems changed librarians' ability to review acquisition decisions, at least for monographs. In 2003, a Mellon Foundation-funded study by the Tri-College Library Consortium (Bryn Mawr, Haverford and Swarthmore Colleges) done in conjunction with the Council on Library and Information Resources (CLIR) found that approximately 75 per cent of the items in the three libraries' collections had circulated no more than once in the past ten years. In addition, about 40 per cent of the items in the collections overlapped (i.e. they were held on more than one campus). About half of these overlapping items had not circulated in the past 11 years.

Galvin and Kent referred to the book budget in the academic world as 'the most sacred of sacred cows' and pointed out:

The hard facts are that research libraries invest very substantial funds to purchase books and journals that are rarely, or never, called for as well as equally large sums to construct and maintain buildings designed to make accessible quickly titles that are no longer either useful to or sought by their clientele. (Galvin and Kent, 1977)

The findings of Galvin and Kent and the Tri-College Library Consortium were undoubtedly distressing to librarians, who typically based their selections in print-dominant libraries on their experience and training to correlate the literature of various genres and disciplines with their users' needs and interests in those fields. Librarians were considered the 'experts' at building collections and their purchasing decisions went largely untested and unquestioned. At the same time, librarians and their sponsoring organisations prided themselves on the size of their print collections, adding shelving at the expense of user space and building additions, new libraries, and high-density storage facilities to house their print library collections.

The older print focused model looked at the collection from the point of view of the collection as a resource. Now, as shown by statistics from the Association of Research Libraries (ARL), the collection is increasingly a service and data collection is guided toward user outcomes.

Assessment and evaluation in the electronic library environment

Networked electronic resources and services for library patrons have become more pervasive, easier to use, and have increased in number and variety since the days when patrons used predominantly print resources. As libraries devote increasingly large proportions of their materials budgets to networked electronic resources, ARL recognised the changed landscape and launched its New Measures Initiative to measure and evaluate usage of networked electronic resources.

According to ARL (2008: 22–3), expenditures for electronic resources among the major research libraries in North America exceeded \$400 million during 2005–06; after including hardware and software and other operational costs, the figure increases to half a billion. This total represents close to 45 per cent of the library materials budget. The need to evaluate the return on the investment made on electronic resources was pressing. Through ARL's E-Metrics initiative (ARL, 2002), efforts were made to define the variables that would be useful to track. In addition, an experimental data collection is currently underway in the form of a supplementary statistics collection of data on searches, sessions and downloads. Evaluation in general is an activity that has increased in libraries over the last decade and it has formalised itself in a number of different ways, ranging from experiments and testbed applications to training, conferences and the emergence of a thriving community of practice (see www.libraryassessment.org).

Documenting contemporary interest in assessment, Wright and White (2007) examined the current state of library assessment activities in ARL libraries. One of the first steps by ARL libraries in developing an assessment programme to learn about their customers was to administer a user survey. The authors found the top five assessment methods currently used by libraries to be statistics gathering, a suggestion box, web usability testing, user-interface usability, and surveys developed outside of the library. Locally-designed user satisfaction surveys used to be more frequent, but have been replaced by the internationally implemented LibQUAL+® survey.

Increasingly, libraries are creating assessment departments, units or positions primarily responsible for assessment activities. Most of these positions or units were created after 2005. Interestingly, assessment results are wide-ranging, and are often specific to the particular library, for example, library facilities, hours, changes to the website, etc. The

identified need for an infrastructure of assessment and a community of practice surrounding it is gradually being realised (DeFranco et al., 2007; Moreleli-Cacouris, 2007).

Usage measurement methods and accountability expectations have also greatly increased. Libraries have new tools for collecting usage data about how users are using digital resources. Both vendor-supplied statistics and locally-generated data through portal websites or gateways document in new ways how patrons are interacting with the library, something that was impossible in the print world. The refinement of e-metrics by librarians to evaluate the use of library resources and services has matured into a set of standardised tools and shared understandings about the value of the metrics for making data-driven, managerial decisions in libraries. E-metrics are applied to a number of library resources and service domains, some as census counts and others as samples. There are four dimensions to e-metrics:

- externally supplied e-metric usage data;
- locally captured e-metric usage data;
- full census data representing all of the usage of networked electronic resources;
- randomly or non-randomly sampled data purporting to represent all usage.

These form a helpful taxonomy to organise methods of collecting e-metric data and to discuss levels of trust by librarians.

Using these four dimensions, usage data can be organised into four categories or types of data collection processes:

- census counts:
 - externally generated, vendor usage data;
 - locally or internally generated usage data;
- sample counts:
 - externally generated, web survey data;
 - internally generated, web survey, usage data.

These four categories of collecting data sum up most of the library-initiated approaches to the evaluation of digital libraries, and define the assessment culture or evaluation environment in which librarians approach the evaluation of digital resources.

Externally-generated, vendor supplied data

Protocols for census counts include the statistics of usage of networked electronic resources collected by external vendors conforming to codes of practice, such as COUNTER, and standards-based expressions of them, such as SUSHI, a standardised transfer protocol for COUNTER-compliant statistics. As documented on its website (www.projectcounter.org), COUNTER proposes a code of practice and protocols covering the recording and exchange of online usage data so that librarians will have a better understanding of how their subscription information is being used, and publishers have standard data on how their products are being accessed. The constantly updated codes of practice recommend that vendors produce and the library use reports containing such variables as the 'number of successful full-text article requests by month and journal', 'turnaways by month and journal', 'total searches and sessions by month and database', in addition to other reports (see www.projectcounter.org/code_practice.html). The SUSHI standard (NISO Z39.93-2007) has three supporting XML schemas posted to the National Information Standards Organization (NISO) website, which also hosts retrieval envelopes for the conforming XML-formatted COUNTER reports (www.niso.org/schemas/sushi/). These data are analysed by libraries, either by moving the data into electronic resource management systems (ERMs) or by creating spreadsheets. The purpose of the analysis is often to generate cost-per-use data. Although the calculation is simple, collecting meaningful cost data from the complex bundling offered by vendors is not trivial.

COUNTER is a tremendous step forward, but not the total solution. From 2004 to 2007, the Institute of Museum and Library Services funded a project entitled 'Maximizing library investments in digital collections through better data gathering and analysis'. As part of the study, three research teams studied different types of usage data for electronic resources in order to develop a cost-benefit model to help librarians 'determine how best to capture, analyze and interpret usage data for their electronic resources' (Baker and Read, 2008). For their part, Baker and Read surveyed librarians at academic libraries to determine the level of effort required to process the COUNTER data, how the data are used, and which data are the most meaningful. Their findings suggest that librarians still wrestle with inconsistent data from non COUNTER-compliant reports, but surprisingly also from the reports of COUNTER-compliant vendors. The process of data analysis

takes time. The most common reason for analysing the data was to make subscription decisions and to meet reporting requirements.

Census data counts like COUNTER are consistent with the tradition of collecting circulation data for print resources. They focus on the collection and not the patron; they result in decisions that create higher frequency counts but not necessarily better patron-centred services; and they measure benefit usually through some sort of cost-per-use metric, not on the value to the patron.

J. C. Bertot, C. R. McClure and D. M. Davis have been pursuing a research agenda to assess outcomes in the networked electronic environment (Bertot and McClure, 2003; Bertot and Davis, 2005; Snead et al., 2005). The approach developed for the Florida Electronic Library looks at functionality, usability and accessibility, and combines a number of iterative methods to assess outcomes. Functionality is defined as a measure of whether the digital library works as intended. Usability assesses how users interact with the programme. Accessibility measures how well the systems permit equal access for patrons with disabilities (Snead et al., 2005). This project has focused on large state digital electronic resource collections, an important target for outcomes assessment. Its strength is that it considers how the digital library (or electronic resources) serves the community as a whole. Part of the evaluation includes usage data from the resources.

The MESUR project (Bollen et al., 2007) seeks to employ usage data to expand the possibilities of scholarly assessment. The purpose is to generate a model of the scholarly communication process involving usage, citation and bibliographic data. It will create a reference set and generate a wider range of usage-based metrics than we presently use, with guidelines for their application. MESUR identifies the current datasets, such as harvestable usage statistics for scholarly journals (COUNTER and SUSHI); the Interoperable Repository Statistics Project (*irs.eprints.org*), which defines usage data for OAI-PMH-compliant repositories; and CiteBase (*www.citebase.org*) which collects citation data (Bollen et al., 2007). The deliverables from this project are a survey and model of the scholarly communication process, a large-scale reference data set for the investigation of viable usage-based metrics, an examination of the clusters of practices found in this data set, and finally the definition and validation of usage-based metrics of scholarly impact.

There are a myriad of business intelligence tools to process COUNTER data. Lakos (2007) gives a useful survey of the state of business intelligence software wedded to the integrated library system and the electronic resource management system. He lists COGNOS,

Voyager Analyzer, SirsiDynix' Directors Station, SirsiDynix' Normative Data Project for Librarians, EBSCO's effort to combine WebFeat's SMART with its A-to-Z management service, Serials Solutions' Overlap Analysis and their 360 Counter, the Ontario Council of University Libraries Scholars Portal Statistics framework, ScholarlyStats from MPS Solutions, and Blackwell's Collection Management Services. The sheer volatility and variety of these statistical reporting and analysis mechanisms speaks to great interest and few standards.

In general, the census data supplied by vendors external to the library can be useful for cost-use studies, although Conyers and Dalton (2007) provide evidence that this analysis is more difficult than it appears. There are nascent efforts to relate such data to analyse user behaviour and motivation. To become user-applicable, the data must typically be combined with locally-generated censuses of web logs or other user data. COUNTER data, though useful for making decisions regarding the collection or cost, are truly revealing regarding users' needs when combined with local data.

Internal locally-developed data

White and Kamal (2006: 129) provide a useful literature survey regarding the collection of data pertaining to the usage of networked resources at the local library level. The collection of such data can be through locally-developed census counts generated from click-through scripts, rewriting proxy server logs, virtual private networks (VPNs), or OpenURL server logs, as well as other methods. White and Kamal also present some creative models of the network infrastructure necessary to collect such data locally, including ERMs (White and Kamal, 2006: 99), VPNs (White and Kamal, 2006: 108), and rewriting proxy servers (White and Kamal, 2006: 109). Unlike external vendor-supplied data, these local data can be mapped against authenticated users or internet protocol addresses to determine usage by local demographics such as client group, school or discipline. Library websites are routinely evaluated by web server logs and web traffic analysis software. ERM applications can tie cost to usage, while enhancing access to e-journals. Interlibrary loan usage can be counted by OCLC's VDX product, Atlas Systems' ILLiad system (now licensed exclusively by OCLC) and Infotrieve's Ariel system. Finally, the online public access catalogue can collect usage data including search statements and success rates. Stemper and Jaguszewski (2003) point out that 'local use data allows us to compare usage across publishers and disciplines'. They conclude that 'it may be useful to occasionally compare

local statistics with vendor statistics to understand usage in more depth', as 'both local and vendor usage data have their own strengths and weaknesses ... Both have their place in the digital library's suite of quantitative evaluation measures' (Stemper and Jaguszewski, 2003).

Transaction logs capture all local usage, yet because of the simplicity of the IP and HTTP protocol elements, they are not particularly useful. If the logs can be tied to a session, that is, one person searching over a period of time, they become more informative. The interaction within the electronic resource is unavailable to the locally collected data, but commensurable counts can be generated across disparate resources. Log files are especially attractive for closed environments, like digital libraries, OhioLINK (www.ohiolink.edu) and OCUL Scholar's Portal (www.ocul.on.ca), and they have relevance to any gateway server, through which requests to e-journal vendors must pass.

In a review of transaction log file analysis and web log analysis, Jamali et al. (2005) note that there are advantages and disadvantages to the technique and that researchers have taken both sides. The advantages include: log file data are collected automatically, data are collected unobtrusively, the data are good for longitudinal analysis, and the data are based on a census not sampling. Log analysis can provide data for the evaluation of digital library performance while providing useful data about information-seeking behaviour.

The disadvantages include the difficulty of differentiating user performance from system performance. It is difficult to identify users, and IP address alone is not sufficient; sessions are hard to determine and many researchers assume that a session lasts 30 minutes; caching proxy servers may thin out the data; and activity by spiders and other crawlers should be segregated in the data. With log file analysis we do not know why the user did what he or she did. There are some problems with caching effects on estimates of volume of use. There are several levels of caching: browser cache, local cache and large regional caching. If the IP address is treated as a user, then proxy servers, firewalls with network address translation, and VPNs bias the logs, as multi-user activity comes from one IP.

Deep log analysis (DLA – see Chapter 7 and Nicholas et al., 2005) enriches web log data with user demographic data, drawing from a user database or online questionnaires. As log files provide little explanation of behaviour, deep log analysis follows up with a survey or with interviews. DLA was developed by the CIBER (www.ucl.ac.uk/ciber/). DLA technique is employed with OhioLINK and is part of the MaxData project, described elsewhere in this chapter. The technique attempts to provide methods for obtaining good-quality usage data through transaction logs, and in this method items used,

viewed or requested are counted as use. Unlike COUNTER, countable *use* includes such outputs as a list of journals, a table of contents page, and the full-text article. COUNTER defines *use* as the download, view, print or e-mail of a full-text journal article. Sessions are terminated after ten minutes of inactivity and there is only one session counted per day.

DLA defines site penetration as the number of items viewed during a particular visit. Time online or page view time is estimated. Users are identified by IP address. Using a reverse domain name system (DNS) lookup, sub-network names are mapped to academic departments. On-campus and off-campus users are differentiated. The user outcomes are difficult to analyse, but one can attempt to track returnees by IP, as a measure of user satisfaction, among other motivations. Assigning of dynamic IPs by internet service providers makes the off-campus returnee data even more speculative (Nicholas et al., 2005). The usage to user effort is well placed, but there appear to be a number of problems, particularly when based upon IP. However, this project seems to have disaggregating user data as one of its goals.

One of the best examples of locally-developed census collection is by Zucca (2008) and the Penn Library Data Farm. In this service-oriented data-collection tool, information is pulled from the online catalogue, acquisitions, circulation, electronic resource management systems, OpenURL link resolvers, interlibrary loan data, web service logs and rewriting proxy server logs, bringing together resources, services and the data they produce when patrons use them. The basic concept is the desire to capture library-related events or interactions in a data warehouse. Digital libraries could fit into this environment very easily. Using these data, Zucca can track resources, people and locations to create a management information framework. A sample report from this system may produce, for example, a view of 'Faculty use of electronic journals by school by journal topic'. Despite its thoroughness and census aspects, this framework would be difficult to replicate elsewhere, and therefore does not produce benchmarked data, but does successfully link usage data to users. This system is particularly useful for libraries assigning library resource and service costs to specific user groups. The possibility of extending this system through an XML schema of standardised event descriptors is under consideration.

Standardised web surveys externally collected

Though usage data from national web surveys rely mostly on perceptions and are subject to limitations of recall and second-hand reporting of user

behaviour, libraries have often asked usage questions in a variety of surveys. In particular, through the well-respected, internationally practised web survey protocol known as LibQUAL+®, users are asked to report their opinions of library service quality. This extensively tested and validated web-based survey helps libraries assess and improve library services. The survey instrument measures library users' minimum, perceived and desired levels of service quality across three dimensions: information control, affect of service, and library as place. Since 2000, over 1,000 libraries and 1 million users have participated in the survey. The survey is marketed locally but administered externally, and the data are collected, analysed and reported externally (www.libqual.org).

Supplemented by comparable data from peer institutions, libraries can focus on the usage related to certain resources. For example, users are being asked how often they use library resources on premises, how often they use resources through library electronic gateways, and how often they use internet gateways such as Google, Yahoo, etc. Clearly the proportion of usage reported that is library-related (both gateways and physical premises) is much less than usage reported for internet gateways such as Google, Yahoo, etc. Using LibQUAL+® data provided by 295,355 of the participants who completed the LibQUAL+® survey in 2003, 2004 and 2005, Thompson et al. (2007) sought to address three research questions. First, in terms of usage of onsite library information resources versus non-library information gateways such as Google, what differences, if any, have occurred over time in the use by (a) undergraduates, (b) graduate students/postgraduates, and (c) faculty? Second, in terms of usage of onsite library information resources versus non-library information gateways such as Google, what differences, if any, have occurred across international regions in the use by (a) undergraduates, (b) graduate students/postgraduates, and (c) faculty? Third, in terms of frequency of onsite and internet gateway usage, what differences, if any, are there in perceptions of library service quality across four user types ('nonusers', 'traditionalists', 'web techies' and 'voracious users'). The results shed light on information use trends and patterns around the world (Thompson et al., 2007).

Internally collected sample counts

Most sample counts are user studies, but are not linked to systematically collected usage counts, nor are the results comparable with peer institutions. Tenopir (2003), updated by Rowlands (2007), has surveyed user studies. One web survey technique that attempts to address sampling and comparability issues is MINES for Libraries®

(*minesforlibraries.org*). The primary difference between the MINES for Libraries® approach and many of the other web-based user surveys recounted in the work of Tenopir and Rowlands is the emphasis on usage. Although user demographic information is collected, this web survey is really a usage survey rather than a user survey. The respondent must choose the web-based networked electronic resource in order to be presented with the survey, thus preventing memory or impression management errors. Once the survey is completed, the respondent's browser is forwarded to the desired networked electronic resource. This approach is based on the random moments sampling technique. Each survey period is at least two hours per month, so each survey period in itself is only a snapshot or picture of usage. Because the survey periods are randomly chosen over the course of a year and result in at least 24 hours of surveying, the total of the survey periods represents a random sample, and inferences about the population are statistically valid with a 95 per cent confidence level and a low standard error (e.g. less than 2 per cent) (Franklin and Plum, 2006).

The MINES for Libraries® survey is most accurate when the library information architecture makes it possible to survey all usage of networked electronic resources during the sample period (an assessment infrastructure), and when response is mandatory. Because the survey is a usage and a user survey, there is a real possibility of repeated surveying of the same patron. One means of reducing the inconvenience to patrons of repeated surveying is to record the values chosen in the initial survey of the patron's usage of electronic resources, and invisibly resubmit those values to the patron for subsequent use with a different networked electronic resource during the sample period. Users' demographics do not change during a session and an examination of the MINES for Libraries® data collected to date shows that repeat users rarely change their purpose of use within a session.

Although MINES for Libraries® has been used at about 50 college and university libraries, (Franklin and Plum, 2006, 2008) it has not achieved wide acceptance, possibly for two reasons: (a) it is implemented locally, and therefore requires local IT support; (b) it is a sample, albeit a well-tested representative random sample, hence it does not have the same level of trust as census counts. To address the IT issues, the MINES for Libraries® protocol is being applied to commonly accepted gateway technologies, such as EZproxy. MINES for Libraries® user data have been successfully linked to EZproxy usage log files to give a fuller picture of the usage of the user, and why resources are being used. Once this linkage is made available to all libraries running EZproxy, then the IT

support issue will become much less constraining. ARL is providing analysis reports to institutions implementing the MINES for Libraries® survey. It is exploring options for scaling the application by automating the collection of data and minimising the reliance on local IT support. To date, the most scalable implementation of MINES for Libraries® is the study implemented at the OCUL Scholar's Portal where, through a central gateway, 16 different institutions participated in the study (Kyrillidou et al., 2006).

How are the data being used?

Cost-benefit analyses

One use of these assessment data is to look at the benefits, as measured in usage volume and the decreased costs of digital materials. Three ground-breaking cost-benefit analysis studies occurred between 2002 and 2004. The first, conducted at Drexel University and reported by Montgomery and King (2002), determined that, while not directly comparable, the total costs (subscription and operational) of electronic journals calculated on a cost-per-use basis were \$1.85 per use, compared with \$17.50 per use for print journals. These calculations were based on article views and downloads for electronic journals and four years of reshelving counts for print journals. Electronic journal use was also much higher than the print journal use measured.

A second study, performed by Obst (2003) at the Medical Branch Library of the University Library in Muenster, Germany, only considered subscription costs. Obst's study also determined considerably lower unit costs for electronic journal use (€3.47) than print journal use (€18.68). Consistent with Montgomery and King's findings, users accessed the electronic versions of journals much more frequently than the print versions. The Muenster study also found significant differences in unit costs by publisher.

A third study, published by the CLIR (Schonfeld et al., 2004), considered the non-subscription costs of current print journals, print journal back files, and electronic journals. This study was interesting in that it attempted to project cost over the estimated total lifespan for periodicals. Again, the authors concluded that, 'other things being equal, an electronic collection should achieve lower non-subscription costs than a print collection'.

Cost and use data are relatively easy to compile for digital resources. With most vendor-supplied and transaction-based usage counts, digital content usage data are based on total usage for a given year, not a sample. The data collected to date indicate that the cost per use of an article in an electronic journal is fairly inexpensive. The more often that digital content is used, the lower the unit cost, and the resulting increase in perceived value to the user population reflected by increased use does not incur additional cost. Therefore, offering digital content encourages the development of library services such as marketing, instruction and liaison outreach. Moving to digital content also nurtures the development of new technology systems to leverage already committed expenses, such as OpenURL, web usability studies, and electronic resource management systems.

Consortia like OhioLINK and OCUL mount commercially licensed digital content locally and calculate cost per use data by title, by publisher, and by member library to determine which electronic resources merit continuation of their licensing and operational costs. Individual libraries measure total use of individual electronic resources and packages and calculate unit costs, usually based on vendor-supplied usage data, to decide which titles and packages have a sufficiently low unit cost to warrant continued investment.

Many libraries now calculate the annual unit costs for electronic books, electronic journals and databases and use that information to inform collection development decisions. Unit cost analysis can help to determine whether publishers' cost increases are justified by measuring whether usage is increasing faster than the costs are inflating. While unit cost data should not be the sole determinant in buying decisions, such data can be used to explain to faculty and students why a title or package may not be a good investment at their university. Unit cost data also help to standardise different publishers, vendors and products so that titles and packages can be evaluated effectively.

As librarians at the University of Montana reported in 2004:

For the first time, the Collection Development Team was able to review the networked resources collection from an overview of quantitative data and relate it to the collection development policy ... At the same time, national-level initiatives to work with vendors to standardize vendor-supplied data provide the future opportunity to further expand the analysis to include how users actually use these resources once they are accessed. (Samson et al., 2004)

Less than a decade into its development, measuring the value and impact of digital content already provides librarians with more useful and accurate data related to collections use than was possible in the print library environment.

Yet these cost-use data are just the start. Cost-use data can be related to the size of the potential population, for example, the biology department of a university. High usage relative to the user population is also factored into cost-usage analysis. Similarly, the more data that can be brought into the analysis about the user, the more informative it is about user behaviour and user outcomes.

Linking usage data to users

One of the more exciting presentations at NISO's 2007 conference 'Understanding the data around us: gathering and analyzing usage data' was the presentation by McDonald (2007), entitled 'Usage statistics and information behaviors: understanding user behavior with quantitative indicators'. In it he made a case for using usage data to understand the impact of networked electronic resources on patrons. He proposed a series of information behaviours, for which he developed a logical outcome that could be measured with usage data. Although the focus was on testing information behaviours, libraries can use a similar methodology to test hypotheses about their users.

The future of information usage collection is linking to user outcomes. Recent refinement of web-based survey techniques promises to provide librarians with even more representative and reliable samples of digital content use. Web-based surveys can be used to measure user satisfaction and digital collection usage that is not vendor-supplied, such as locally-mounted digital collections and services and locally-hosted open access journals. Surveying at the campus router level or at the proxy rewriter provides a comprehensive sampling plan that is able to survey all electronic services users, regardless of their point of entry.

These assessment efforts allow librarians to better understand digital content usage, make more informed collection development decisions and to better justify collections management choices. However, librarians are now able to more accurately determine who is using specific digital content. Knowing the locations where networked services are being used (e.g. the majority of faculty prefer to work on campus, but not in the library) enables librarians to plan user support services accordingly. Determining purpose of use permits academic librarians to

identify which electronic resources contribute most to their institutions' primary missions of instruction/education/unfunded research, funded research, patient care, public service, and other institutional activities.

The response of practising librarians to these advances in digital library evaluation has not been universally enthusiastic. Some librarians have responded positively to management information related to collection development decisions, and to the possibilities of relating usage to users, but others see usage data as threatening the values and practices that librarians cultivated in the print environment. For many years, librarians carefully reviewed published titles and selected library materials that, in their opinion, matched local users' needs and interests. Some libraries deliberately continued to build strong focused collections on particular topics that 'characterised' their library collection, sometimes on the basis of a donor's gift and often without regard for whether those collections were used sufficiently to justify their ongoing costs. Evaluation of library collections usage in the networked resource environment has, in effect, exposed some of these traditional collection development practices as failing to successfully match library collections with local users' needs and interests. The digital library environment has provided librarians with the tools to evaluate their purchasing decisions and to review usage of specific elements of the networked electronic resources, including electronic books, electronic journals and databases. Over time, usage data will become more standardised and linkages to users outcomes more routine. Service will improve and patrons will be more empowered to achieve their outcomes.

What do library metrics and librarian expectations mean for the assessment of digital libraries?

Librarians have a framework of expectations about the evaluation of digital libraries, which is part of their assessment culture. Institutional repositories are digital libraries supporting the university mission, and are the focus of much attention by librarians, particularly because of the documented non-participation by faculty (Davis and Connolly, 2007). Through interviews with faculty, Davis and Connolly sought to determine factors for faculty non-participation in Cornell University's installation of DSpace. Although a number of the reasons given by faculty for not using this particular digital library are specific to institutional repositories

(such as the learning curve, redundancy with discipline repositories, copyright, intellectual property, etc.), the emphases on service and scholarly purpose were clear. Only one faculty member focused on perceived failures of the software, access, searching, retrieval or preservation. For most faculty members, the primary deterrents were the perception of problems with the service aspects of the institutional repository, and its alignment with the faculty mission of publication.

Institutional repositories are a specialised type of digital library, and the evaluation of institutional repository literature is mature, with numerous checklists, indicators for success and evaluation criteria (Center for Research Libraries, 2007; JCDL, 2006; Proudman, 2007; OARiNZ, 2007; Westell, 2006). For example, Mary Westell's 'Institutional repositories: proposed indicators of success' (2006) lists a clearly defined mandate, integration into institutional planning, ongoing funding model, relationship to digitisation centres, interoperability, measurement, promotion, and preservation strategy. Measurement includes usage data, but more importantly also includes measuring the contributions of repository content to scholarship. Demonstrating linkages between repository content and scholarship is an incentive to faculty to contribute content, but also demonstrates the alignment of the repository with the university mission.

The evaluation of institutional repositories, and the assessment culture created by the various approaches to evaluation and assessment lead to some generalisations about how libraries and librarians approach the evaluation of digital resources. The following observations about how librarians approach evaluation are not strict rules but common points taken from the evaluative efforts to date:

- *Evaluation should be based on standardised usage data collected within the local digital library, and benchmarked to national norms:* The librarian expects to see standardised data across different digital libraries and digital library platforms representing usage. The librarian understands the implications of COUNTER data, and would want to integrate digital library usage data into library data analysis systems. Evaluative systems for digital libraries producing idiosyncratic or non-standard, locally-collected data are not going to be trusted. For example, the NSDL 'Evaluating digital libraries: A user-friendly guide' (Reeves et al., 2006) lists a traditional set of evaluation tools to inform decision-making.
- *Census not samples:* The librarian is familiar with many user survey methods, but because they are based on samples, they are usually case

studies in specific institutions. Ad hoc changes can be made to library facilities and websites based on user surveys, as described in Spec Kit 303, but for systematic evaluation, librarians will expect census data, or a national sampling plan such as that used by LibQUAL+®. In general, usage data are much more trusted than user survey data, unless the user surveys are very well done with a solid sampling plan, and are shown to be reliable and valid.

- *Web usability is an important, formative evaluation process rather than summative for digital libraries:* Librarians are confronted with dozens of different interfaces and systems. The best libraries regularly run web usability studies on their websites, and make changes to their web presentation based on the data. Web usability protocols are one of the methods that have had the greatest impact on libraries over the last years (Hiller et al., 2008). This is becoming an ongoing improvement process for all library systems.
- *Content is still the king:* Digital libraries represent, in some manner, a return to the ownership of collected materials, organised and made available by the library, now in digital form, but with the new digital assessment tools to collect and analyse usage data. The digital library based on usage data will live or die by its content. As Rowlands (2007) notes, researchers are not especially accomplished searchers and use various strategies including browsing to work around their own lack of familiarity with the digital collection search interface, or its limitations. Nonetheless, scholarly digital collections are perceived as increasing productivity. The content of the digital library is more important than its interface.
- *Usage data and user outcomes for service improvement:* There is an expectation that usage data will be linked to user outcomes. Librarians are protective of the privacy rights of their patrons, so if this linkage is accomplished through authentication, librarians will insist that the data are stripped of ID information after aggregation. The purpose of linking user outcomes to usage data is to inform decision-making. Assessment results in some action that improves service for patrons.
- *Service competition for users:* Digital libraries must compete with other library resources and services. The stronger the case the digital library can make for affecting user outcomes, the better it will be accepted in the library vision of service to its users. Valid usage data related to clear user outcomes tells a convincing story to the librarian, who must also assess numerous other services, often by sophisticated

cost-use or cost-benefit analysis. These services all compete for funding and attention in the library world. To some extent, the bottom-line evaluation of the digital library in the library world is how it competes in the marketplace of resources and services available to users.

Bibliography

- Association of Research Libraries (2002) 'Measures for electronic resources (E-Metrics)', available at: <http://www.arl.org/bm~doc/e-metrics.pdf.zip> (accessed 8 August 2008).
- Association of Research Libraries (2008) 'ARL Statistics 2005-06', available at: <http://www.arl.org/bm~doc/arlstats06.pdf> (accessed 28 July 2008).
- Baker, G. and Read, E. J. (2008) 'Vendor-supplied usage data for electronic resources: a survey of academic libraries', *Learned Publishing* 21(1): 48-57.
- Bertot, J. C. and McClure, C. M. (2003) 'Outcomes assessment in the networked environment: Research questions, issues, considerations, and moving forward', *Library Trends* 51(4): 590-613.
- Bertot, J. C. and Davis, D. M. (2005) *Planning and Evaluating Library Networked Services and Resources*, Westport, CT: Libraries Unlimited.
- Bollen, J., Rodriguez, M. A. and Van de Sompel, H. (2007) 'MESUR: Usage-based metrics of scholarly impact', in *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries Vancouver, BC, 17-23 June*, New York: ACM Press, p. 474.
- Conyers, A. and Dalton, P. (2007) 'Electronic resource measurement: Linking research to practice', *Library Quarterly* 77(4): 463-70.
- Center for Research Libraries (2007) *Trustworthy Repositories Audit and Certification (TRAC): Criteria and Checklist*, Chicago, IL: Center for Research Libraries and OCLC.
- Davis, P. M. and Connolly, M. J. L. (2007) 'Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace', *D-Lib Magazine* 12(3/4), available at: <http://www.dlib.org/dlib/march07/davis/03davis.html> (accessed 8 August 2008).
- DeFranco, F., Hiller, S., Janicke Hinchliffe, L., Justh, K., Kyrellidou, M. and Self, J. (eds) (2007) *Library Assessment Conference: Building*

- Effective, Sustainable and Practical Assessment*, Washington, DC: ARL.
- Franklin, B. and Plum, T. (2006) 'Successful web survey methodologies for Measuring the Impact of Networked Electronic Services (MINES for Libraries)', *IFLA Journal* 32(1): 28–40.
- Franklin, B. and Plum, T. (2008) 'Assessing the value and impact of digital content', *Journal of Library Administration* 48(1): 1–57.
- Galvin, T. and Kent, A. (1977) 'Use of a university library collection', *Library Journal* 102(20): 2317–20.
- Hiller, S, Kyrillidou, M. and Self, J. (2008) 'When evidence isn't enough: Organizational factors that influence effective, sustainable and practical assessment', *Performance Measurement and Metrics* 9(3): 223–30.
- Jamali, H. R., Nicholas, D. and Huntington, P. (2005) 'Use and users of scholarly e-journals: a review of log analysis studies', *Aslib Proceedings* 57(6): 554–71.
- JCDL (2006) 'Digital curation and trusted repositories: seeking success', available at: <http://sils.unc.edu/events/2006jcdl/digitalcuration/> (accessed 8 August 2008).
- Kyrillidou, M., Olshen T., Franklin, B. and Plum, T. (2006) *MINES for Libraries™: Measuring the Impact of Networked Electronic Services and the Ontario Council of University Libraries' Scholar Portal, Final report, January 26*, Washington, DC: ARL.
- Lakos, A. (2007) 'Evidence-based library management: The leadership challenge', *portal: Libraries and the Academy* 7(4): 431–50.
- McDonald, J. (2007) 'Usage statistics and information behaviors: understanding user behavior with quantitative indicators', paper presented at NISO 2007: Understanding the Data Around Us: Gathering and Analyzing Usage Data, Dallas, TX, 1–2 November, available at: <http://niso.kavi.com/news/events/niso/past/usage07/usage07mcdonald.pdf> (accessed 8 August 2008).
- Montgomery, C. H. and King, D. W. (2002) 'Comparing library and user related costs of print and electronic journal collections: a first step towards a comprehensive analysis', *D-Lib Magazine* 8(10), available at: <http://www.dlib.org/dlib/october02/montgomery/10montgomery.html> (accessed 8 August 2008).
- Moreleli-Cacouris, M. (ed.) (2007) *Library Assessment Conference Papers, 13–15 June 2005, Thessaloniki, Greece*, Washington, DC: ARL.
- Nicholas, D., Huntington, P., Jamali, H. and Tenopir, C. (2005) 'What deep log analysis tells us about the impact of big deals: case study OhioLINK', *Journal of Documentation* 62(4): 482–508.

- OARiNZ (2007) 'Evaluation criteria', available at: <http://www.oarinz.ac.nz/oarinzwiki/index.php?n=OARiNZ.EvaluationCriteria> (accessed 8 August 2008).
- Obst, O. (2003) 'Patterns and cost of printed and online journal usage', *Health Information and Libraries Journal* 20(1): 22–32.
- Proudman, V. (2007) 'The population of repositories', in K. Weenink, L. Waaijers and K. van Godtsenhoven (eds), *A DRIVER's Guide to European Repositories*, Amsterdam: AUP, pp. 49–101.
- Reeves, T., Apedoe, X. and Woo, Y. (2006) 'Evaluating digital libraries: A user-friendly guide (revised version)', available at: <http://www.dpc.ucar.edu/projects/evalbook/index.html> (accessed 28 November 2008).
- Rowlands, I. (2007) 'Electronic journals and user behavior: A review of recent research', *Library and Information Science Research* 29(3): 369–96.
- Samson, S., Derry, S. and Eggleston, H. (2004) 'Networked resources, assessment and collection development', *The Journal of Academic Librarianship* 30(6): 476–81.
- Schonfeld, R. C., King, D. W., Okerson, A. and Gifford Fenton, E. (2004) *The Nonsubscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats*. Washington, DC: CLIR.
- Snead, J. T., Bertot, J. C., Jaeger, P. T. and McClure, C. R. (2005) 'Developing multi-method, iterative, and user-centered evaluation strategies for digital libraries: functionality, usability, and accessibility', paper presented at the Annual Meeting of the American Society for Information Science and Technology, Charlotte, NC, 28 October to 2 November, available at: <http://eprints.rclis.org/archive/00005268/> (accessed 8 August 2008).
- Stemper, J. A. and Jaguszewski, J. M. (2003) 'Usage statistics for electronic journals: an analysis of local and vendor counts', *Collection Management* 28(4): 3–22.
- Tenopir, C. (2003) *Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies*. Washington, DC: CLIR.
- Thompson, B., Kyrillidou, M. and Cook, C. (2008) 'On-premises library use versus Google-like information gateway usage patterns: a LibQUAL+® study', *portal: Libraries and the Academy* 7(4): 463–80.
- Westell, M. (2006) 'Institutional repositories in Canada – institutional repositories: proposed indicators of success', *Library Hi Tech* 24(2): 211–26.

- White, A. and Kamal, E. D. (2006) *E-Metrics for Library and Information Professionals: How to Use Data for Managing and Evaluating Electronic Resources Collections*, New York: Neal-Shuman.
- Wright, S. and White, L. S. (2007) *Library Assessment: ARL Spec Kit 303*, Washington, DC: ARL.
- Zucca, J. (2008) 'Building frameworks of organizational intelligence', paper presented at the Library Assessment Forum, University of Pennsylvania, Philadelphia, PA, 11 January, available at: <http://www.libqual.org/documents/admin/zuccaarl08.ppt> (accessed 8 August 2008).