The Economics and Usage of Digital Library Collections

by Kate Thomes, Head, Bevier Engineering Library, University of Pittsburgh

The conference on "The Economics and Usage of Digital Library Collections," held in Ann Arbor, Michigan, 23-24 March 2000, and sponsored by the Program for Research on the Information Economy and the University Library at the University of Michigan, provided an opportunity for an international group of librarians, publishers, and economists to speak on the economics and usage of digital collections. The conference also marked the conclusion of the PEAK project, a four-year study of digital collection pricing models and user behavior sponsored by the University of Michigan. The PEAK report was one of 16 presentations on a variety of research studies and practical applications of pricing and distribution models, economic analyses, user behavior studies, and the impact of digital libraries on traditional library operations. The full text of most of the papers presented is available, along with more information about the speakers and projects, from the conference website at <http://www.si.umich.edu/PEAK-2000/>. This article summarizes the main themes that emerged from the conference.

Distribution of Innovation

Academic Servers

Innovation in scholarly communication is forging ahead in a variety of arenas now that digitization and the World Wide Web allow informal distribution of information. Discipline-based servers dedicated to research fields (e.g., Paul Ginsparg's physics server--arXiv.org--or the RePEc economics server) or specific concepts (e.g., Neil Sloane's integer sequence server) are becoming essential platforms for scholars. These are sources for preprint and peer-reviewed articles, primary research, and an interactive exchange of ideas and even materials.

Traditional Publishers

Commercial ventures, such as Kinko's and Amazon.com, are also having a significant impact on the publication and distribution system that was once solely the domain of libraries and traditional publishing houses. The trend toward publisher-determined aggregations of journals with full-text access to articles and the added value in several traditional abstracting and indexing services that provide the ability to link from a citation to the full text of an article are examples of publishers' efforts to redesign their services to take maximum advantage of digital technology.

Library & Publisher Cooperative Initiatives

SPARC and JSTOR are examples of cooperation between libraries and publishers to maintain the values of academe in the publishing and distribution of, and access to, scholarly resources in the new digital landscape.

User Behavior in Digital Library Collections

Obsolescence of Journals?

There is growing awareness that users want collections of articles rather than collections of journals. Research findings from the PEAK project support this conclusion. In the PEAK study, 12 colleges and universities received access to all the content from 1,200 Elsevier Science titles using a variety of pricing and subscription models. Data from the study show the 80/20 rule in action: 80% of the use came from 20% of the articles. The PEAK study also found that library patrons used articles from nonsubscribed journals to a higher degree than anticipated.

JSTOR data also support this idea, showing that a few articles are used repeatedly while others have never been used.

Usage Patterns
Paul Kantor (Rutgers University School of Communication, Information, and Library Studies) reported on user behavior from the Columbia Online Books Project, a longitudinal study of use of online scholarly monographs. Among other things, this study revealed a variety of usage patterns for online books. For example, some users viewed the material in order (A, B, C, D), others jumped around (B, D, A, C) but viewed each section only once, others jumped back and forth (A, C, A, D, B, C). The pattern of use varied by the type of book (textbook, tradebook, scholarly book). This study also found that digitized books received about three times the use of their print counterparts, in terms of the number of downloads compared to the number of times an item was checked out from the library.

Users Highly Sensitive to Obstacles
Andrew Odlyzko (AT&T Labs), Clifford Lynch (Coalition for Networked Information), Bob Gazzale (PEAK), and others discussed another aspect of user behavior: even the slightest barriers to access discourage users. These barriers, or "costs," include a multitude of physical limitations in traditional libraries (e.g., limited hours, misshelved or lost material, confusing classification systems) and new barriers for digital resources (e.g., long connect times, multiple search interfaces to learn, use of passwords, and reluctance to pay for material that used to be "free"). It was also pointed out that, to a growing degree, material that is not in the online catalog or another online format "ceases to exist." Users expect the resources they need to be easily available online and are less willing to track down print materials than they were in the past.

Research Opportunities
Data collection capabilities from digital collections are far more sophisticated, reliable, and precise than from print collections, providing abundant opportunities for research on usage and user behavior. User behavior data from the Columbia experiment is an example of the detail now available to researchers. Wendy Lougee of the PEAK project expressed the need to develop new metrics to measure a variety of aspects of digital libraries; she stressed the value of engaging economists in this discussion.

Impact of Digital Libraries on Traditional Library Operations
In response to the digitization movement, libraries are changing the ways they manage budgeting, staffing, and roles or work functions.

Expensive Added Value
Bruce Kingma (State University of New York at Albany) presented results of a study performed for the Canadian Institute for Historical Microproduction, which revealed that initial costs for digitizing were significantly higher than for microfiche reproduction but that use of the digitized version greatly exceeded the use of the other formats. This indicates that digitization may be cost effective over time, after the high start-up costs have been absorbed.

Shifting Workloads
Drexel University is an example of an early adopter of digital collections. Carol Montgomery described Drexel’s move from a collection of 100 e-journals and 1,850 print journals in 1998 to a collection of 5,000 e-journals and 953 print journals in 2000. She described this transition’s impact on staffing, shifting workloads, and new job functions. Administration, management, and computer network infrastructure all saw increases in responsibility. Technical services functions--including e-journal acquisitions, cataloging, and catalog maintenance--all required increased staffing. In addition to obvious reductions in staff to check-in, claim, and bind printed journals, fewer staff were needed for reshelving and stack maintenance. Montgomery believes that building and maintaining a digital library collection is far more complex than doing the same for a print collection due to the price and license negotiations that are required.

At one point, Montgomery stated, "It's hard to get rid of print!" and presented a short list of print-weeding strategies. Missing from her list, but not for long, was the suggestion from JSTOR's Kevin Guthrie to send back-runs of journals to JSTOR for archival digitization.

Economics of Scholarly Publishing
All the presentations acknowledged to some degree that the economics of scholarly communication are changing rapidly.

Serials Crisis Overview
Don King, co-author with Carole Tenopir of *Towards Electronic Journals* (Washington: Special Libraries Association, spring 2000), presented findings of research on scholarly journal publishing from 1960 to 1995. His presentation provided some historical context and helped explain how prices for journals--science journals in particular--increased so dramatically over the last 15 years.

Mary Case of ARL described early efforts by libraries to respond to these price increases, including journal cancellations, reduced monograph acquisitions, improved document delivery networks, and consortia. In 1989, ARL commissioned an economics consulting firm to analyze scholarly publishing trends. Their report pointed to the lack of competition in this market as a major contributing factor to spiraling costs. SPARC was created in 1997 by ARL to inject competition into the scholarly publishing market by facilitating the start-up of low-cost/high-quality academic publications. Case also described SPARC's history, partnerships, and current projects. SPARC has demonstrated remarkable impact already with a variety of new publications and outreach programs that have raised understanding of these complex issues across the nation.

Market Analysis
Mark McCabe, Georgia Institute of Technology, provided an economist's view of the market for scholarly communication. In explaining the serials crisis, he described library demand for scholarly publications as "inelastic," i.e., demand for the material is not affected by its price. Publication by academics is viewed as so essential (to both communication and certification processes) that no cost is considered too high and libraries are expected to pay whatever is necessary to obtain the material.

Further, library demand for scholarly publications creates a unique market environment. In the paper he presented, McCabe says, "Although most STM (science, technology and medicine) journals are highly differentiated even within sub-disciplines, cost per journal citation is minimized across a broad field of study, subject to a budget constraint, and the result is a demand for portfolios of titles. In other words, unlike most markets involving differentiated products, it is not appropriate to model demand as a discrete choice process. Rather, the typical library attempts to provide access to as many STM journals as possible through a combination of subscriptions and inter-library exchanges."

As commercial publishers analyzed the demand structure of libraries, they came to understand the profitability of acquiring more titles through mergers. As a few publishers gained greater market power they were able to increase the price of individual titles within the portfolios they had to offer.

McCabe concludes that these three factors--inelasticity of demand, library acquisition by portfolio within broad subject areas, and publisher mergers--all contributed to a higher rate of journal price inflation over the last several years than can be explained by the improved quality or increased publishing costs of the journals themselves.

Changing Sources of Revenue
Science magazine's Michael Spinella pointed out that the traditional revenue source for magazines--paid

advertising--is not viable in the digital environment; people do not run across ads serendipitously online as they do in print. This is causing major rethinking of the economics of online publishing for several publishers. Both Spinella and Karen Hunter of Elsevier Science said their organizations are trying to develop flexible subscription and delivery options to accommodate the needs of a variety of individuals, colleges, and universities.

Conclusions

In wrapping up the conference, Jeff MacKie-Mason of the University of Michigan restated what several of the speakers had noted: electronic access is increasing the use of materials but we do not yet fully understand that use. Who is using the materials? From where is the use coming? What value is the use delivering?

Users are going to the World Wide Web for their information, in many cases bypassing the library. What does this mean for the role of the library in the future? Traditionally, libraries have maintained the authoritative record of scholarship and made that record broadly available. Will this continue to be the case?

Libraries must communicate the actual costs of digitizing collections to a wider audience and digitization must be understood not as a cost saver but as a value-added service.

Another theme raised is that major technological change and adaptation to new technologies takes five to ten years. We are probably in about year three in the move from print to digital collections. Changes in the next few years will, in all likelihood, continue to be breathtaking.

Digital Initiatives

arXiv.org is a preprint physics server created by Paul Ginsparg and supported by the U.S. National Science Foundation, the U.S. Department of Energy, and the Los Alamos National Laboratory.

Columbia University Digital Library Collections Online Books Project developed a comprehensive evaluation methodology to be applied to the University's pilot project in online books. The Project, funded by a grant from The Andrew W. Mellon Foundation, ran from 1995 through 1999.

JSTOR, Journal Storage, is an independent nonprofit organization established to help the scholarly community take advantage of advances in information technology by digitizing the complete backfiles of core scholarly journals.

PEAK, Pricing Electronic Access to Knowledge, was a project sponsored by the University Library and the Program for Research on the Information Economy at the University of Michigan. The project, which ran from 1996 through 1999, was a large-scale trial of production-quality digital library services hosted by a nonprofit intermediary, and a field experiment on the economics and usage of digital access to scholarly communications.

RePEc, Research Papers in Economics, is a series of interoperating electronic archives of academic economics papers.

Sloane's Online Encyclopedia of Integer Sequences is a searchable database of number sequences arranged in lexicographic order.

SPARC, the Scholarly Publishing & Academic Resources Coalition, is an alliance of research institutions, libraries, and organizations that fosters expanded competition in scholarly communication.

Table of Contents for Issue 210 | Other Scholarly Communication Articles | Other Networked Information Articles | Other Statistics and Measurement Articles

ARL Bimonthly Report Home ARL Home

Maintained by: ARL Web Administrator Site Design Consultant: Chris Webster **Last Modified:** August 5, 2002