W(h)ither Health Science Libraries

Dynamics and Effects of Digital Materials Use on the Future Roles of Health Science Libraries

February 7, 2002

Thomas C. Rindfleisch
Director Emeritus, Lane Medical Library
Stanford School of Medicine

Contact: tcr@stanford.edu

Introduction

Since the middle 1990s, when world-wide-web tools and content seriously began their rise, on-line digital information has transformed many aspects of our lives, including research, health care, education, government, commerce, the law, and entertainment. From the perspective of the health sciences — research, clinical care, education, and patient information — the effects of digital information have been among the most dramatic [1]. These are persistent, on-going changes and have not been affected by the demise of less substantive applications and services with the bursting of the "dot-com" bubble in the last half of 2000. These technology changes, along with other pressures, in turn are having a profound effect on health science libraries — the traditional locus of information management, access, and use in health settings.

In recent years most institutions have moved aggressively toward providing scholarly information (especially periodicals) in digital forms [2] and some have sought broader scale digital integration [3]. Independent of the digital information revolution, health science libraries have been under mounting pressure because of traditional resource competition within academic medical schools — involving principally money, space, and personnel slots [4, 5]. The rise in availability, demand, and use of digital materials have only exacerbated this pressure on libraries — by increasing costs for acquisitions [6], by making users less dependent on the physical collections and study spaces libraries offer because they can access digital information from their workplaces, and by progressively disintermediating and changing the role of libraries as organizers and facilitators of access to information.

The study reported here addresses a central part of this changing scene —the relative usage rates of on-line and print information, the dynamics of usage changes over time, and the long-term roles of libraries in biomedicine. This study has been done in the context of Stanford University's Lane Medical Library, a mid-sized health science library supporting Stanford's medical school, hospitals and clinics, and the university as a whole. One of the motivations for this study has been the need to upgrade the Lane Library and medical school classrooms, which were built in 1959 and have not been upgraded in any substantial manner for over 40 years. The underlying questions relate to, what kind of education and library facilities does a front-line medical school need for the coming decades?

Methodology

Since about 1996, Lane Medical Library has made a concerted effort to make more and more information available to its user community in digital form. Lane now facilitates access to about 1,000

biomedical and clinical journals, 50 textbooks, and many other databases in full-text digital form[1]. As is typical of modern health science libraries, the use of journals dominates over the use of books and other resources by at least an order of magnitude. For this reason, we emphasized journals and, in particular, high impact titles in our study. We wanted to measure usage rates as a function of (a) full-text on-line availability (along with hardcopy access), and (b) the age of the volume involved. Digital access is (in principle) fairly easy to track based on web site activity. We are fortunate to have close collaborative ties with HighWire Press [7], which provided detailed on-line usage information to us (as permitted by particular publishers). The digital access measure we used was based on viewing journal tables of contents and downloads of full-text articles (in HTML or PDF forms), and did not count simple viewing of titles and abstracts available through the National Library of Medicine's (NLM's) PubMed resource [8].

The core of our study was to analyze on-line and hardcopy usage, with the measures noted above, for two groups of high-impact scientific (as opposed to clinical) journals. These journals were chosen from among the most frequently used titles within the Stanford community, as determined by ISI Science Citation Index[2] records for publications and reference citations in articles by Stanford scientists and overall reshelving statistics (see Figure 1). The first study group included 5 titles that have been available on-line in full-text form, as well as in hardcopy form, since 1998 (see Table 1). The second group includes 5 titles that have been available only in hardcopy form to the Stanford community (see Table 2).

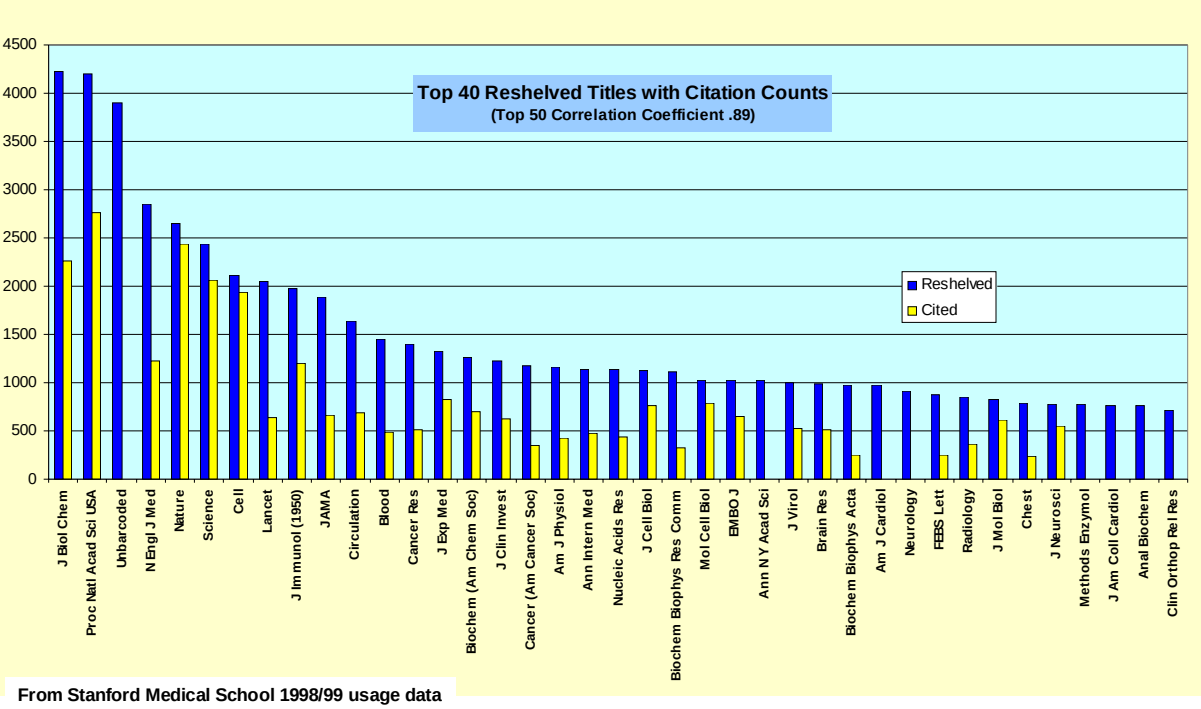

Figure 1: Most frequently used and cited journals for Stanford Medical School

[1] Lane Library's digital offerings can be reviewed at http://lane.stanford.edu/lane/reference/indexej.html.

[2] See http://www.isinet.com/isi/products/citation/sci/index.html.

Title	ISSN	Sponsor/Publisher
Journal of Biological Chemistry	0021-9258	American Society for Biochemistry and Molecular Biology
Journal of Cell Biology	0021-9525	Rockefeller University Press
Journal of Experimental Medicine	0022-1007	Rockefeller University Press
Proceedings National Academy of Sciences USA	0027-8424	National Academy of Sciences
Science	0036-8075	American Association for the Advancement of Science

Table 1: Journal titles available to the Stanford community in both digital and hardcopy forms

Title	ISSN	Sponsor/Publisher
Brain Research	0006-8993	Elsevier Science
Cancer Research	0008-5472	American Association for Cancer Research
Cell	0092-8674	Cell Press
Journal of the American College of Cardiology	0735-1097	American College of Cardiology
Journal of Immunology	0022-1767	American Association of Immunologists

Table 2: Journal titles available to the Stanford community only in hardcopy form during this study

We used journal circulation and reshelving data as a measure of library hardcopy usage. At Lane, study areas and copy machine areas are distant from the stacks and users are encouraged to leave volumes on study tables or in utility shelves near copy machine alcoves rather than doing their own reshelving. Lane began collecting detailed reshelving data for its collection in the early-mid 1990s, when collection volumes were progressively tagged with unique bar codes. Local modifications to the CARL library automation system, in use at Lane since 1986, allow recording of reshelving date and time for each unique volume. In this way we have detailed usage data from in-house use as well as from circulation use by volume. This surrogate usage measure for library hardcopy volumes is something of an underestimate of the corresponding total institutional use we seek to measure in that it does not capture use of unbound library issues (bar codes are only assigned to volumes as cumulated runs of recent issues are bound) or use of personal and departmental copies of journals. It is hard to quantitate how much this would add to the "hardcopy usage" measure, and is, of course, title dependent since expensive subscriptions will be rarer in personal and departmental settings than more affordable subscriptions.

For each calendar month (1/1999 - 6/2000) and title, we recorded the total digital use (number of retrievals of tables of contents or full text of an article) for on-line titles independent of the date of the issue containing the article. We also retrieved the reshelving records for combined in-house and circulation use for each calendar month (8/1997 - 6/2000) and title. We grouped the reshelving records for each title into two parts for each month: (1) reshelving count of volumes <= 3 years old at the date of reshelving, and 2) reshelving count of volumes > 3 years old at the date of reshelving. We then plotted these results for each title and performed a least squares fit of the reshelving data over the period 6/1998 - 6/2000. This period corresponds to that during which digital information had become widely available and the library had focused efforts on announcing the availability of digital information and

had offered routine training in its use to faculty, staff, residents, and students. This was also a period after the announcement of free access to PubMed by the NLM.

Results

Full detailed data relating to this study will be made available after completing and publishing the results. For this preliminary summary, we offer typical plots for one of the on-line titles (*PNAS* — see Figure 2) and one of the hardcopy-only titles (*Cell* — see Figure 3).

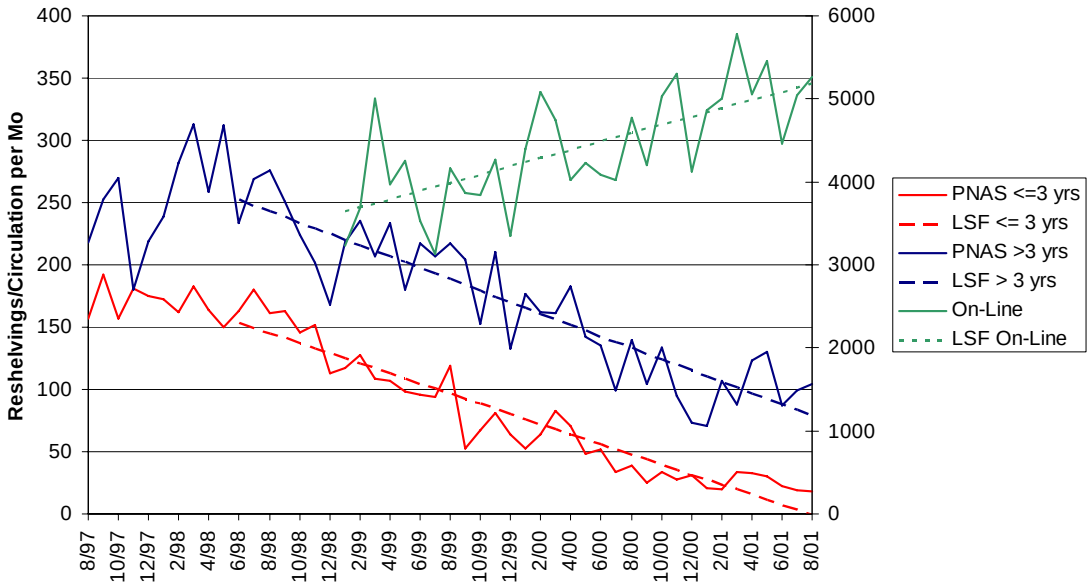

Figure 2: Reshelving statistics and digital use for the *Proceedings of the National Academy of Science*

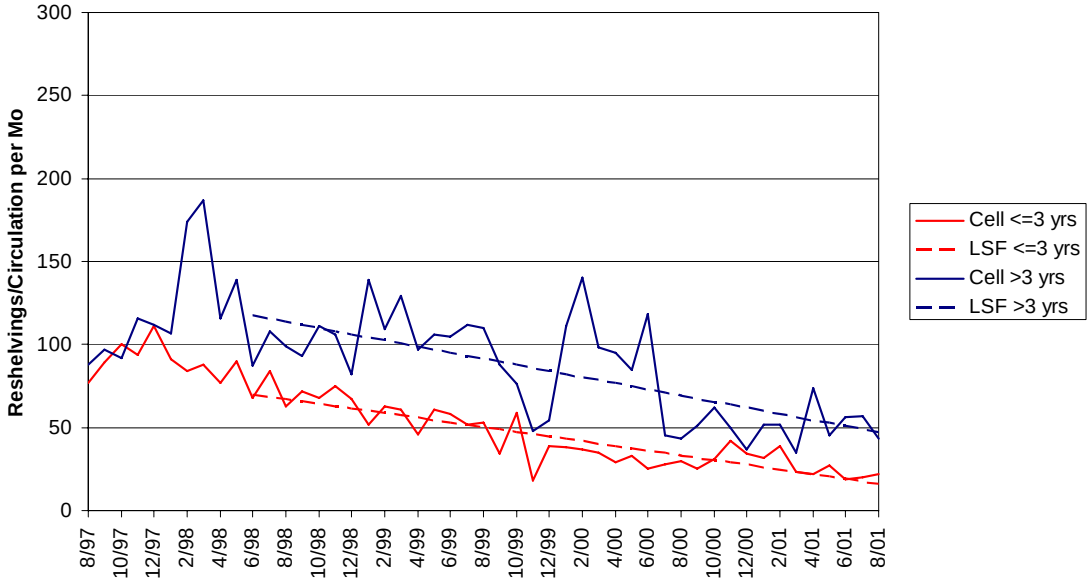

Figure 3: Reshelving statistics for *Cell*

For both of these titles, independent of the month-to-month variation in usage, there is a distinctly apparent drop-off in use. This behavior is characteristic of that for all of the titles in the two groups. In fact if we summarize the slope data for the least squares fits to usage for 6/1998 - 6/2000, we find the following results (see Table 3).

	Loss in Reshelving Recent Volumes (per Yr)	**Loss in Reshelving Older Volumes (per Yr)**	**Digital Use/ Hardcopy Use**
On-Line Titles			
J Biol Chem (5/95)	-45%	-16%	21
J Cell Biol (5/97)	-51%	-18%	18
J Exptl Med (1/97)	-42%	-21%	10
PNAS (1/97)	-42%	-19%	15
Science (11/95)	-36%	-10%	46
Average	-43% +/- 4.9%	-17% +/- 3.8%	22 +/- 13
Hardcopy-Only Titles			
Brain Res	-46%	-21%	
Cancer Res	-37%	-8%	
Cell	-31%	-2%	
J Amer Coll Card	-37%	-28%	
J Immunol	-33%	-22%	
Average	-37% +/- 5.2%	-16% +/- 9.6%	

	Change in Reshelving Recent Volumes per Year (w/r 6/98)	Change in Reshelving Older Volumes per Year (w/r 6/98)	Change in Digital Use per Year (w/r 9/00)	Ratio of Digital Use to Hardcopy Use
On-Line Titles				
J Biol Chem (5/95)	-39% (R2 = 0.77)	-15% (R2 = 0.28)	+23% (R2 = 0.68)	21
J Cell Biol (5/97)	-40% (R2 = 0.69)	-14% (R2 = 0.15)	+20% (R2 = 0.16)	23
J Exptl Med (1/97)	-40% (R2 = 0.73)	-18% (R2 = 0.30)	+14% (R2 = 0.24)	12
PNAS (1/97)	-37% (R2 = 0.89)	-22% (R2 = 0.70)	+16% (R2 = 0.48)	21
Science (11/95)	-31% (R2 = 0.72)	-14% (R2 = 0.22)	+32% (R2 = 0.40)	60
Average	-37% +/- 3.4%	-17% +/- 3.1%	+21% +/- 6.3%	27 +/- 17
Hardcopy Titles				
Brain Res	-36% (R2 = 0.74)	-17% (R2 = 0.19)		
Cancer Res	-26% (R2 = 0.52)	-13% (R2 = 0.20)		
Cell	-29% (R2 = 0.76)	-14% (R2 = 0.21)		
J Amer Coll Card	-34% (R2 = 0.59)	-28% (R2 = 0.40)		
J Immunol	-32% (R2 = 0.68)	-19% (R2 = 0.31)	+11% (R2 = 0.05)	20
Average	-31% +/- 3.6%	-18% +/- 5.3%		
All Lane Titles	-25% (R2 = 0.80)	-14% (R2 = 0.48)		
Total sample ave	-34% +/- 5%	-17% +/- 4%		
P-value (T-Test)	0.040	0.621		

Table 3: Summary of reshelving and digital usage data for on-line and hardcopy-only titles.

Overall hardcopy usage trends for all titles subscribed by Lane, are shown in Figure 4. As before the trends for recent and older volumes are shown separately and project losses of 25% and 14% per year respectively.

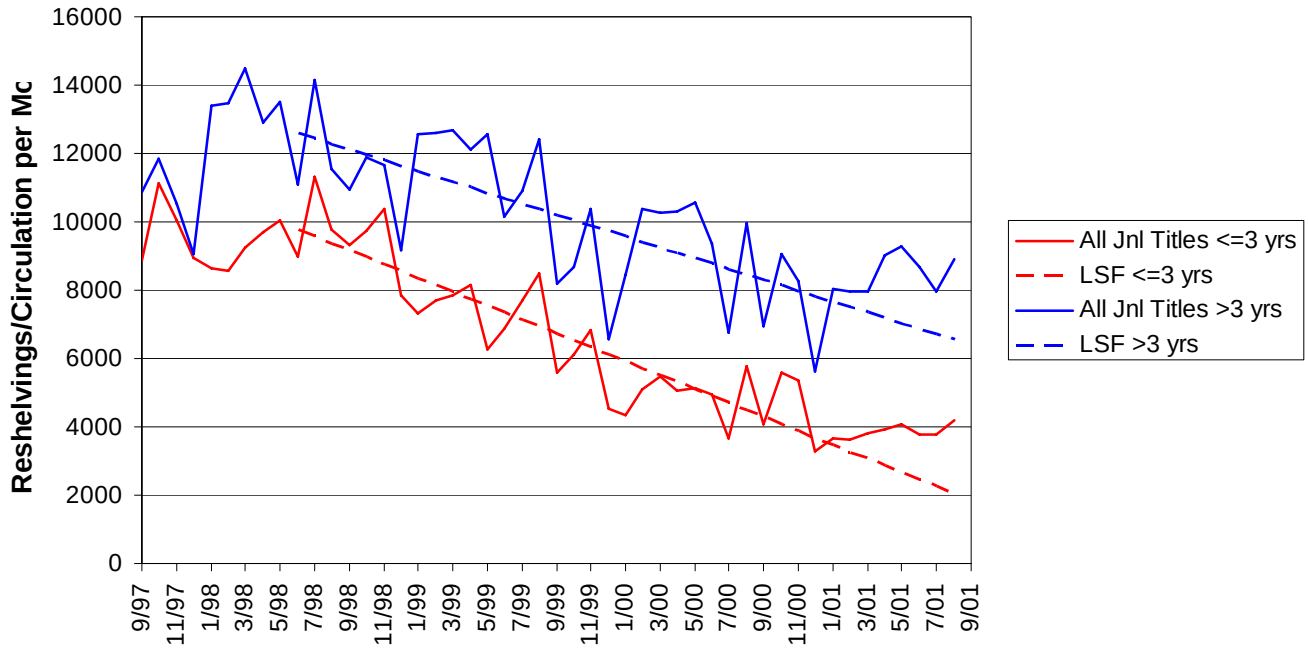

Figure 4: Hardcopy usage trends for all subscribed titles.

Discussion

A number of observations appear relevant and justified concerning these data. First, for all titles the overall use of volumes <= 3 yrs is about the same as the use for volumes > 3 yrs old. This, of course, emphasizes the fact that modern biomedicine is changing so fast that literature use is heavily biased to recent years. Second, based on the HighWire data, digital use is about 20x library hardcopy use, but is fairly flat during 1999-2000 — i.e., there is no apparent dramatic growth in on-line use over that already achieved for these titles. Third, a more surprising result is that library use of hardcopy-only titles seems to be declining at almost the same rate as use of on-line/hardcopy titles (-37% per year as compared to -43% per year). From anecdotal discussions with Stanford faculty, it seems clear that a "laziness factor" is taking over — people are getting so used to getting what they want on-line from their offices that if a title is not on-line they make do with the MEDLINE abstract or give up retrieving the article all together (Lane does not do document delivery). Access to older volumes is not declining as much as to recent volumes (-16/-17% per year vs -37/-43% per year) but still the "writing seems to be on the wall". Given that on-line titles currently go back to about 1995 and the vast majority of our literature use occurs within the first 10 years of publication, this means that by about 2005 we will have a very good collection of on-line materials and any title that resists coming on-line will get lost in the transition. In the meantime, we may suffer from poor scholarship by depending on abstracts (rather than full text) for articles not readily accessible on-line.

This drop-off in use of library-based hardcopy literature is paralleled by a drop-off in use of the library itself (see Figure 5). At Lane, student study and multi-media areas are separate from the main library, so the primary reason for coming to the core library is for hardcopy collection access, public workstation

access, information service consultation, or study table/carrel use. Lane keeps an accurate count of patrons entering the library and the total number of entries per month has been declining continuously for the past several years. This effect is correlated with a first-ever drop in library copier use since early 2000.

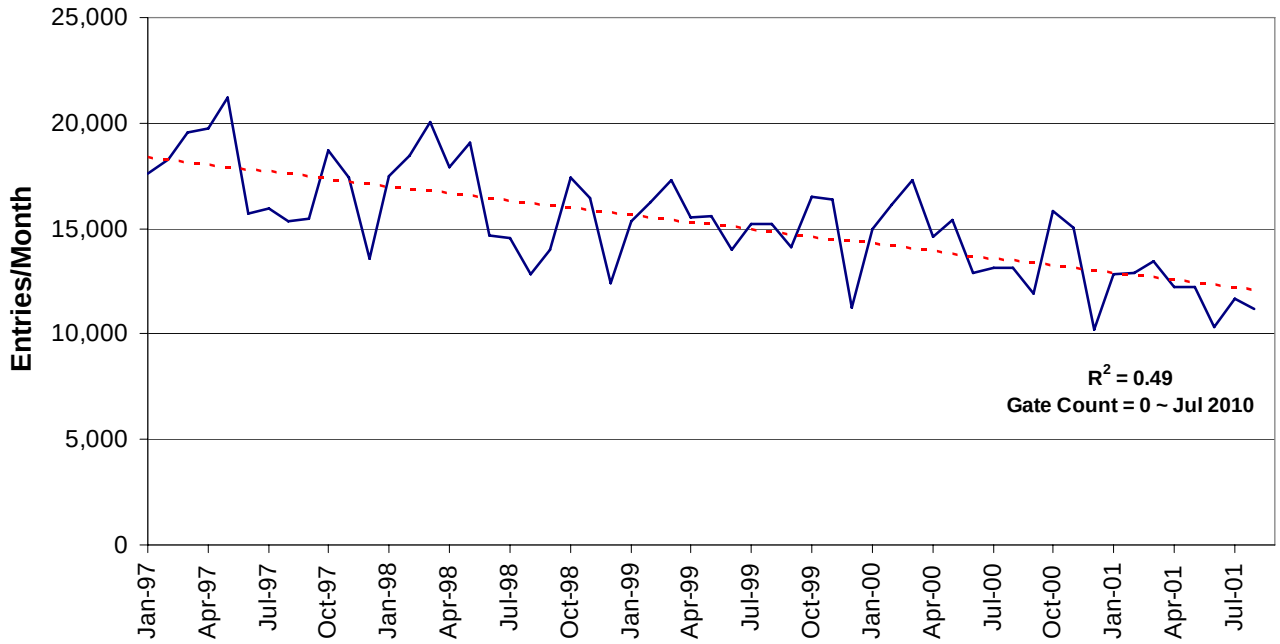

Figure 5: Lane Library gate count trend

Making any predictions about future roles for libraries is, of course, a very complex topic that depends on what institution we are talking about and where it fits in the hierarchy of resource libraries (e.g., national library, medical school library, hospital library, or general library), what discipline we are talking about (e.g., genomic research, history of medicine, or the law), and the scope of the role the library plays in the institution's infrastructure (e.g., collection management, digital licensing, study space, computing and networking resources, instructional media and classrooms, teaching, and other services).

These data suggest strongly that the role of health science libraries is changing dramatically with the rise in availability and use of on-line full-text information. As the emphasis on physical collections diminishes, we can expect libraries to retain their role in providing physical space for study, classrooms, a locus of informatics work, and student socializing. With the growing complexity of the types of information available on-line and the need for well integrated computing and networking access, we can expect libraries to play a central role as a source of user help for setting up their computing and networking tools and for accessing, searching, and using information. Finally, I would emphasize the growing role of libraries in areas of technological innovation — as producers of accessible content (ala HighWire Press); as institutional licensing representatives and brokers of information access; as archivists of information in digital forms (including institutional historical information); as collectors, managers, and integrators of diverse content for target user communities; and as operators and support resources for digital media facilities in health care settings (e.g., for students, residents, patients, and faculty).

This move to digital forms of information is not without its growing pains, including the need to evolve workable economic models for publishers, institutions, and consumers of biomedical information. Librarians are already playing substantial roles in helping shape institutional policies and practice in this area. Also, digital forms of information have historically not been stable for archival access. The evolution of incompatible magnetic tape, magnetic disk, and optical forms of storage over recent decades does not bode well for being able easily to recover information 100 years from now, much less 10 years from now (see for example a recent OCLC/RLG report on *Attributes of a Trusted Digital Repository* [9]). Central libraries will undoubtedly continue play a lead role in developing the necessary technologies and in implementing them operationally.

Even though the use of modern biomedical literature is heavily biased toward current periodical publications, many users (e.g., writers of review articles, students trying to gain a perspective on modern biomedical science, and serious medical historians) will continue to require access to older documents of many sorts. How are we to facilitate this access, especially in an era of increasing pressures on academic space? Many institutions have undertaken a "caching" approach wherein only the most recent and heavily used volumes are kept on-site, and old or rarely used volumes are kept off-site. These are accessed either by users requesting copies of particular articles, visiting study rooms at the remote site(s), or by requesting delivery of physical volumes for use on-site. These approaches are currently the most cost-effective from institution administration points of view, but ultimately impede use of the historical record. MEDLINE, for the most part, does not index articles before 1962 and only started storing abstracts in the 1970s. Most library catalog systems have only meager "finding aids" in that little more than titles, authors, and related meta-data are cataloged. Full table-of-contents and article-level information (ultimately full text) would help greatly in locating relevant information. The most desirable visionary solution would appear to be to digitize the entire historical record so that it could be stored more economically (and irredundantly) and could be searched more effectively by users to locate and conveniently retrieve relevant materials. Some efforts have begun in this direction (see for example the JSTOR [10] scholarly journal archive project), but converting the entire historical record to digital form has formidable technical and economic obstacles. Scanning old documents often requires debinding (and damaging) the volume and the cost is still prohibitive (current estimates for scanning and optical character recognition run at least $1-2 per page). The ultimate solution to achieve this vision for easy access to historical information would have to be mounted as a national (or international) effort to amortize the costs and to take advantage of the benefits appropriately. At the very least, we should assure that new additions to the archive from current times forward are stored in digital (as well as backup hardcopy) forms.

Acknowledgments:

It is a pleasure to acknowledge the contributions of Dick Miller (Head of Technical Services for Lane Library), Randy Woelfel (Systems Programmer who extracted the reshelving and gate count data from Lane database records), and John Sack and Michael Newman (who facilitated access to digital usage data from HighWire Press).

Bibliography:

1. Hersh, W.R. and T.C. Rindfleisch, *Electronic publishing of scholarly communication in the biomedical sciences.* J Am Med Inform Assoc, 2000. **7**(3): p. 324-5.
2. DLF, *Digital Library Federation*. 2002.
3. Kvaerndrup, H.M., *DEF - Denmark's Electronic Research Library: A Project Changing Concepts, Values And Priorities.* Exploit Interactive, 2000(4).
4. Boyd, S. and A. Herkovic, *Crisis in Scholarly Publishing*. 1999, Stanford University.
5. Roberts, P., *Scholarly Publishing, Peer Review and the Internet.* First Monday, 1999. **4**(4).
6. Chen, F.L., P. Wrynn, and J.L. Rieke, *Electronic journal access: how does it affect the print subscription price?* Bull Med Lib Assoc, 2001. **89**(4): p. 363-371.
7. HighWire, *Stanford University Libraries, HighWire Press*. 2002.
8. PubMed, *National Library of Medicine*. 2002, National Library of Medicine.
9. OCLC/RLG, *Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources*. 2001.
10. JSTOR, *The Scholarly Jounal Archive*. 2001.