

DAIRL

A BIMONTHLY REPORT ON RESEARCH LIBRARY ISSUES AND ACTIONS FROM ARL, CNI, AND SPARC

MANAGING DIGITAL ASSETS IN HIGHER EDUCATION: AN OVERVIEW OF STRATEGIC ISSUES

by Donald J. Waters, Program Officer, Scholarly Communications, The Andrew W. Mellon Foundation

Teaching and research in higher education are increasingly dependent on primary and secondary sources that are available in forms other than traditional paper-based books, journals, and manuscripts. To reflect the diversity of materials, especially in digital form, and their importance to academic pursuits, it is increasingly common to refer collectively to these sources as “assets” of a college or university, or of the academy in general. The notion of “managing digital assets” is also surely meant to allude to the realm of finance and insurance, and to suggest that we are operating in a realm where change is rapid, and where prudence, skill, imagination, courage, and a good deal of risk taking are coins of the realm. However, let me say plainly at the outset that the digital assets that are the subjects of this paper are resources for research and teaching in higher education, and that the aim of academic institutions in managing them is to advance knowledge and improve education. To paraphrase the slogan from the 1992 Clinton campaign: “It’s the scholarship, stupid!” The touchstone question for all the topics covered in this paper must be: How well does this resource, or that system or feature, advance scholarship?

I also want to emphasize, as a corollary principle, that these assets—these resources for research and teaching—are the lifeblood of the academy. I was reminded recently that not every college and university recognizes the centrality of their research and teaching resources in explicit declarations like mission statements or even in practical day-to-day decision making. For purposes of this discussion, I take my cue from the great religious historian, Jaroslav Pelikan, who in 1992 published a very useful book entitled *The Idea of the University*:

A Reexamination. Originally delivered as a series of lectures at Yale University, it is a meditation on and attempt to update John Henry Newman’s classic work on higher education. In his *Reexamination*, Pelikan identified four core and enduring functions of higher education: research; teaching; the dissemination of knowledge through publication; and the preservation of, and access to, the scholarly record in libraries and archives. The latter two functions—dissemination and preservation and access—refer to the life cycle of scholarly resources that are used and produced in teaching and research and are the objects of scholarly communications.

If, as seems evident, scholarly resources in digital form are playing an increasingly important role in scholarly communications, then we must also acknowledge by way of introduction, that the allocation of roles and responsibilities for managing these resources will be equally important and will shape the future of higher education. These roles and responsibilities are in flux and the future is uncertain, but as we plan for the future, we might usefully turn for comfort and lessons to the history of the development of other communications media, such as print, with which we are more familiar. Some of you undoubtedly have read the fascinating work by Adrian Johns entitled *The Nature of the Book: Print and Knowledge in the Making*. One of his primary theses is that print, which we now all but take for granted, did not simply emerge in the conventional form that we now know it. Rather, these conventions were laboriously constructed over time. As Johns tells the story, there was little confidence in the medium at first, which was full of fraudulent uses that were difficult to distinguish from the legitimate ones. The term piracy, in fact,

“seems to have been coined by John Fell, bishop of Oxford, to describe the rapacious practices of [early] London printers and booksellers. It had a technical meaning: a pirate was someone who indulged in the unauthorized reprinting of a title recognized to belong to someone else.” The development of printing, according to Johns, was nothing less than the working out of a complex “taxonomy of practices labeled piratical—from piracy itself, through abridgement, epitomizing, and translation, to plagiarism and libel,” and the allocation of trust and credit to some of these practices and distrust and discredit to others.¹

These manufacturers of credit, as Johns usefully calls them—the men and women of the trade who were in the thick of this battle to define legitimate print practices—were subsumed under the title “stationer.” Stationers were responsible for the art and mystery of printing, much like today’s technologists are responsible for the art and mystery of digital communication, but the tasks for which they were responsible were many and varied, including functions that we would today identify in separate roles such as printer, publisher, editor, reviewer, wholesaler, and bookseller. For print, stationers worked out over time the conventional practices of making books, which in turn came to affect and determine the making of knowledge. And as they did, print became sustained as a viable economic enterprise and the elaborate, complex, infrastructure of producing a book eventually became invisible to all but the practitioners in the trade. Similarly, specialists in libraries, publishing houses, universities, and in upstart new businesses over the past several decades have been busy sorting out the various tasks associated with scholarly communications in a digital world. Eventually, these roles will settle out into a stable ecology of interaction—but the effort promises to continue for the foreseeable future.

As this sorting out occurs, where will the action be in the coming decade? I suggest that we look for clues, first, to where the action has been in the recent past. One of the largest, most visible, and fastest growing investments in scholarly communications by colleges, universities and scholarly publishers over the last decade has been in the shift to electronic journals and databases, and in a corresponding shift from purchasing to licensing. I would suggest that the implications of these shifts will be critical in the coming decade—and perhaps in some unexpected ways. To tease out these implications, I first want to review and critique the current debate about scholarly publishing. Then I will turn to acknowledge the big elephants lurking about in our midst—namely, Google, Yahoo, and Microsoft—and highlight how questions and issues might shift in light of what we are learning about the role of search engines.

The “Crisis” in Scholarly Communications

The flash point of the recent debate about scholarly publishing has been the phenomenon of steadily increasing prices. In the early 1990s, a Stanford economist named Roger Noll and his colleague, Edward Steinmueller, identified a basic structural effect in scholarly publishing that causes journal prices inexorably to rise. Increasing specialization of knowledge and the interest of scholars in publishing in the “best” journal in their specialty leads to a proliferation of journals. Specialty journals have a narrower readership than established journals of more general scope and may siphon readers from them. As readership falls, journals then require increasingly higher prices simply to cover basic costs.² This pattern of specialization, journal proliferation, and price increases has been evident for more than the last 50 years with the fantastic expansion of fields in science, technology, and medicine since World War II, and it helps explain why the “crisis in scholarly communications,” especially in scientific, technical, and medical fields, has been a perennial problem, the focus of several national commissions, and vigorous, ongoing discussion.³

With the introduction of digital versions, which are easier to use for many purposes and much in demand, the crisis has not abated, as many, including the Mellon Foundation, hypothesized it would. Instead, it has in some ways worsened. Prices continue to rise at rates significantly above inflation.⁴ Meanwhile, publishers seek both to protect their revenue streams against cannibalization effects, as digital files have leaked out and circulated to those who did or might have purchased a print version, and to help finance the retooling of their technical systems. Moreover, a classic example has played out in the academy of the pricing distortions that can arise from the so-called principal/agent problem, that is, when agents—provosts and librarians—act in place of the principal consumers—the scholars. Rather than resisting enormous annual price increases, those agents largely agreed, especially during the booming ‘90s, to finance them. A provost once justified such payments to me by saying: “It is my job to feed the hearts and minds of the faculty and staff. For healthy hearts, I must provide extraordinary annual increases in budget for health care; for their minds, I must provide similar increases in the serials budgets.”

Now that resources are squeezed, attention has focused on the increasingly deleterious effects of these previous financial agreements while yet still neglecting the basic underlying structural problem that Noll and Steinmueller identified. On many campuses, administrators have mobilized faculty, calling them to account as editors and authors of expensive journals and

involving them more deeply as principal consumers in ongoing purchase and cancellation decisions. Big deals have been undone, and bundles unbundled. Researchers and federal regulators have scrutinized publisher consolidation for cause to justify government antitrust action. Publishers, for their part, are increasingly militant about plugging the digital leaks from interlibrary loan, electronic reserves, and other forms of file sharing. In addition, members of the academic community have embarked on a vigorous search for other, alternative business models. This search has resulted in stimulating and sometimes heated debates about the viability of a suite of options under the broad umbrella of “open access.”⁵

In its narrow formulation, open access publishing would disrupt the current system by shifting the burden of generating revenue from the demand side through widespread use of subscriptions, to the supply side “by charging authors or their sponsors for dissemination, or by some kind of institutional subsidy,” making use cost-free.⁶ Such a shift would have the benefit, in theory, of putting the principals—the scholars—back in the economic driver’s seat, and it would have the broad public policy benefit of lowering the economic barriers to reading and using the publications. Discussion of this idea has quickly revealed that in very few disciplines do scholars have sufficient funds from grants and other sources to pay author fees, and that there would be an administrative nightmare if academic institutions that are already financially strapped for funds were to massively reallocate budgets from library acquisitions and other sources to support author fees in any kind of fair and equitable fashion. Publications adhering to this model of open access will undoubtedly continue to be created and survive, but they will probably be limited in number unless and until sources of supply-side revenue can be found other than grant support, and that do not require fundamental administrative and financial overhauls of our institutions.

A broader approach to open access recognizes these practical difficulties and focuses on other ways to lower the barriers to access, such as by encouraging publishers to make articles freely accessible after a limited time during which they exploit subscription revenue,⁷ or by embracing the older call for authors to “self-archive,” that is, to retain rights to make their articles openly accessible in preprint and/or postprint form.⁸ One theory is that if enough authors were self-archiving, then new services could arise to collect, aggregate, evaluate, and present these articles to users. At some tipping point, as yet undetermined, these services might serve to challenge and undermine the economics and inefficiencies of the current system of publication.⁹

The key barrier to a complete transformation following this scenario, however, is an asset management

issue. Does it make sense for systems of knowledge to be built upon the fragile infrastructure of a network of personal Web sites that are subject to personal whims, not to mention the migratory habits of individuals? Institutional repositories might help, especially if they were to collect other related faculty output, such as underlying data and teaching materials.¹⁰ Cheap, easy-to-use-and-manage publishing tools, such as those being developed at some universities, and increasing reliance on sophisticated “recommender” systems to judge relevance and quality after publication rather than before might also help, especially if they are applied in the early stages of rapidly advancing new knowledge, where innovative means of documenting new knowledge are most needed.¹¹ However, in order to justify the necessary and significant costs of such repositories and tool development, colleges and universities must develop compelling rationales for collecting, preserving, and providing access to these kinds of scholarly output. Moreover, these innovations must be invested with features of the current scholarly publishing system that preserve trust in the authenticity of academic work and reliably allocate credit.

Because there is so much at stake for individual faculty in the ways that the current system confers credit and authenticity, it appears that these factors are going to be the hardest to disrupt. Even the self-archived material in physics and related disciplines in Paul Ginsparg’s famous ArXive has not resulted in substantial shifts from traditional forms of publication—at least not yet. Still, efforts to build new models of scholarly communications based on rights to self-archiving, institutional repositories, and innovative publishing tools have gained growing interest, and they remain worth exploring.

Looking beyond the “Crisis”

Important as the serials crisis is, and as pregnant as the discussions about open access alternatives may be, there are even larger forces at play. These are only partially revealed in the system of scholarly publishing, and may even be obscured by a narrow focus on pricing and open access. First, whatever happens with open access, it is not likely to result in a uniform, utopian solution. As Jason Epstein has written, “the global village green will not be paradise. It will be undisciplined, polymorphous and polyglot.”¹² The academic world is and will undoubtedly remain highly pluralistic. Just within the domain of publishing, traditional journal production will remain with us for some time and the shift of those journals to electronic forms of dissemination is likely to continue.

Moreover, our institutions have a lot to gain economically in this transition from print to electronic publishing. Not only do electronic publications provide

CURRENT ISSUES

Continued

greater functionality for teaching and research than those in print, but also as a recent study published by the Council on Library and Information Resources has shown, there is good evidence that the considerable operational costs in libraries of ordering, receiving, processing, shelving, and circulating physical copies can be eliminated by a shift to electronic versions. In the aggregate across institutions, these potential savings may total in the tens of millions annually.¹³ It is worth noting, however, that these are the costs in the print world that represent our system of archiving. Cutting those costs without putting in place reliable preservation archives that are committed to the academic mission remains a problem, and there is, unfortunately, a widely held view among our academic institutions, including libraries and publishers, that savings are theirs to capture and reallocate, and that covering the costs of preserving digital assets for the long term is a responsibility for someone else.

In the face of this reasoning, let me cast the archiving problem in even starker terms. The shift to electronic publication in its current form represents a dramatic, jump-off-the-cliff shift in the academy from owning scholarly output to effectively renting it.¹⁴ The rallying cry about pricing asserts that the academy is giving away its products only to buy them back. However, under current licenses, research and academic libraries do not “buy back” content in the sense of taking local possession of a copy as they did with print. Rather, they use content stored on remote systems controlled by publishers. Moreover, current licenses limit use so that more traditional, mostly regional initiatives for collaborative collection development and resource sharing across institutions are now next to impossible. Instead, the hundreds of thousands of dollars going out the door each year typically buy only a year’s worth of access to the resource and only for members of a single institution. And even if a license recognizes so-called “perpetual access” rights for the material to which an institution subscribes each year, these rights are largely theoretical.¹⁵ Typically, publishers promise to transfer the material on a pile of CDs or tapes, but I am not aware that any such transfer in a perpetual access claim has ever taken place, and it is unlikely that any institution has or will build the capacity to implement such a solution. Not only have prices risen, but the material terms of the licensing deals are transforming—and I would suggest severely weakening—the underlying

infrastructure of resources available for teaching and research.

The negative consequences of electronic publishing that I have outlined were clearly not intended, and they may not yet overwhelm the benefits that the academy has gained, but the unintended consequences could be even worse. Let me suggest a scenario that starts with the observation that many of the largest publishers have now achieved the price increases that they need to survive in an electronic-only world. Their journal business is now declining as a proportion of sales, and unit prices are also declining as they take advantage of the technology and add content. They are moving on to new businesses. The one that has the most promise is offering data-mining services. Such services require large

aggregations of electronic journals, data, and other materials, and promise to help scholars uncover hidden connections and new lines of research. The large publishers have already formed an aggregation of citations in CrossRef,¹⁶ and they could build a similar one for their journal content and other related materials. In a not-for-profit organization, like CrossRef, which manages a cross-publisher aggregation, it would be a short step to build preservation into the mission and for the aggregation to become a publisher-created preservation archive.

This scenario is highly plausible, perhaps more so than promises about the transformative nature of open-access publishing, and the implications are at least threefold: (1) libraries will not own the publications that form the scholarly record; (2) libraries will not own the archive of the scholarly record; and (3) publishers will charge whatever the market can bear for data-mining services because they control all the underlying resources. In other words, if universities and libraries fail to act responsibly and soon in creating archives of electronic journals and other scholarly resources, and publishers act instead, the way will be clear for them to complete a massive transfer of wealth and control over the scholarly record.

The need for action is urgent, and in a licensing regime, the key strategic element for library action is the archives they choose to create now at this key point of transition. Fortunately, there are hopeful signs. In the US, archiving solutions, like LOCKSS and Portico are maturing and successfully recruiting publisher and library participants.¹⁷ The Library of Congress’s National Digital Information Infrastructure and Preservation Program (NDIIPP) has mounted a

...the material terms of the licensing deals are transforming—and I would suggest severely weakening—the underlying infrastructure of resources available for teaching and research.

significant effort on this issue for a broad range of materials involving a large set of institutions.¹⁸ In addition, a growing number of senior officers of our colleges and universities—presidents, provosts, and chief financial officers—are beginning to understand the huge risk to the future of their institutions' core operations caused by the growing dependence on a record of scholarship for which the institution is paying substantial sums but on which it has no real continuing claim. They are beginning to recognize that a preservation infrastructure is needed across the board, not just for electronic journals, and that to build it they must act collectively across traditional institutional and disciplinary boundaries. Against the backdrop of these strategic issues, it is hardly worth mentioning that open access, however it is defined, does not in itself provide an answer to the preservation problem; it simply presents the problem in another form.¹⁹

The Elephants in Our Midst and Their Implications

An even more significant set of strategic issues that has the potential to profoundly and permanently disrupt the patterns of higher education is what Lorcan Dempsey of OCLC calls the "Amazoogle effect."²⁰ It is now well known, and still deplored by some, that Amazon, Google, Yahoo, and other online systems are the first and sometimes only stops for students doing research. Faculty, too, have come to depend increasingly on these services. Amazon has been working closely with publishers for years to make the contents of current publications more accessible and "search friendly." Google too has tried to achieve similar goals, although there is growing evidence that it is clumsy, even inept, in its relationships with publishers and libraries. Google Scholar, which was announced in the fall of 2005, is far from comprehensive in its coverage, but its ability to parse out citations from articles, among other remarkable features, shows how Google can bring exceptional technical expertise to some of the more nuanced and specialized needs of scholars. And this is not the end. As the first anniversary approached of Google's announced plans to launch a large scale retrospective digitization project based in five major research libraries, Yahoo and Microsoft announced their participation in a related effort, led by the Internet Archive, focused on out-of-copyright and public domain materials, and called the Open Content Alliance (OCA).

One of the most common figures of speech that has appeared in public discussions of digitization over the last decade has been the invitation to imagine having the entire Library of Congress available electronically and accessible at the click of mouse. The investment of both Google and the Internet Archive in re-engineering the digitization process and of significantly reducing the

costs so that each could undertake its own initiative means that the vision of digitizing the holdings of our largest research libraries is not only imaginable but may actually be within reach.²¹ These initiatives and the related and competitive projects they stimulate could be incredibly valuable for the public and for the academy in particular. But that Google, Yahoo, and Microsoft are undertaking this effort, not for philanthropic purposes, but for business reasons, means that higher education—at least its library and publishing arms, which are responsible for collecting, preserving, providing access to, and disseminating content of scholarly significance—now has formidable for-profit competitors with considerable resources and their sights set squarely on key parts of the higher education business.

The outcomes are far from certain. The relationships between the work of the Open Content Alliance and that being undertaken in Google's library projects are unclear. Publishers and authors are now suing Google for copyright infringement in the arrangements it has made with libraries. Moreover, in making secret deals, and failing to articulate coherent and collective public interest objectives, the Google 5 libraries may well have squandered a substantial part of the public trust, which they and their institutions have taken decades, even centuries, to earn.²² But let us leave these concerns aside and assume that, one way or another, large-scale digitization of the kind envisioned by Google and its partners and by the Open Content Alliance takes place.

Among the big strategic questions for higher education would be how scholarly communications should be organized in such an environment. These questions have scarcely been identified, much less aired and fully discussed. I am going to leave a number of these issues to one side today and instead highlight several other broad implications of the potentially disruptive influences of Google and OCA on the academy and particularly on the ways that the academy manages and uses its scholarly resources.

The "Processed" Publication

First, I want to draw attention to an idea that Joseph Esposito highlighted a few years ago in a *First Monday* article.²³ For scholars, massive digitization and open access are not ends unto themselves. The central issue is whether scholars can advance knowledge in ways that were not previously possible. Scholars need to make use of digitized and open access materials. Esposito's insight is that at the highest level of generality, what unites our interest in digitization and open access in a digital world is that the material becomes "processable," or subject to computational processing. That is, the growth in the market of readers is not among groups of humans, but of machines, which are programmed to index, manipulate, mine, aggregate, decompose, and

build up scholarly and other forms of content by algorithm. It is this machine “processability” that makes digitized objects and open access materials most valuable to scholars.

Intellectual Property

This brings me to a second point about intellectual property. The temptation is to throw up one’s hands in despair at the massive cost of meticulously clearing the rights of every rights holder in an object that is to be made “processable,” and either to abandon digitization of copyrighted material altogether, or to engage in efforts—also costly but often not accounted for—to stay under the radar of the copyright police. These approaches stand in contrast to a growing set of initiatives, including Mellon-funded initiatives such as JSTOR, ARTstor, Columbia International Affairs Online (CIAO), the American Council of Learned Societies (ACLS) History E-Book Project, the BiblioVault project at the University of Chicago, the Electronic Enlightenment at Oxford University, and New World Records, all of which demonstrate that communities of users and publishers can find ways to create the needed trust and goodwill and agree to overcome the costly barriers of copyright to create highly useful, digitized and “processable” collections of research and educational materials.

For the enterprising and clever, there are countless business opportunities here to be tried and exploited. To pave the way for such entrepreneurial activity and economic growth, intellectual property, and the bundles of rights and duties that surround it, may need to be re-conceptualized, but not in the terms of the already stale and vitriolic debate about copyright and “copy-left.” Rather, there may well be a need and opportunity to learn from Adrian John’s history of print and to recalibrate licenses, intellectual property rights, and even copyright law itself against a richer taxonomy of uses appropriate to digital technologies, many of which may currently be regarded by design or default as “piratical.” Google, for example, seems prepared to respond to the infringement suits by arguing that creating an index is a fair use of copyrighted material.²⁴ This is a clever and enterprising argument and might be plausible if only Google swore off other uses of the copies of copyrighted materials it would make, which it is apparently unwilling to do. Because machine indexing has become such an integral part of the infrastructure today for serving the US constitutional principle of promoting “the progress of science and the useful arts,” legislation that redefines such indexing as legitimate and provides the necessary ancillary protections may well be warranted—and preferable to an ambiguous court decision in a fair use case.

Search

Third, I would highlight the need for new and expanded search and research capabilities as one example of the type of entrepreneurial activity that is needed to build the necessary infrastructure for future scholarly communications. Google’s indexing of full text in its library projects would be generated by optical character recognition (OCR) and could greatly expand and facilitate basic searching and retrieval. Serious thought now needs to be given about ways that Google and other search engines’ technologies could be used to achieve the metasearch and other service objectives we are trying to achieve, sometimes at great expense, in the catalogs of our local systems.²⁵ However, we also need to be thinking imaginatively beyond the local system and traditional library catalogs.²⁶

The sheer volume of digitized material, for example, would require implementation of much more sophisticated indexing, searching, and filtering techniques, including broad application of computational linguistic and related statistical techniques as well as sophisticated techniques for filtering based on markup and thesauri, which would relate results to discipline-based concepts and concerns.²⁷ Above all, there will be growing demand for mechanisms to link search results flexibly across systems in ways that resemble but will be fundamentally different from metasearching across catalogs. To provide a simple example: how easily could one search for related materials in ARTstor, and JSTOR, and, say, Readex NewsBank? Google or Yahoo may be able to respond to a basic demand for cross-searching, but as scholars become more sophisticated in their use of these technologies, their needs will become correspondingly more specialized and discipline-specific in ways that it will likely be unprofitable to address for commercial companies aimed at the mass market. Search and information retrieval is a growth industry not only in the general economy but also for scholarly communications. Solutions that the large search engines cannot supply will have to come from search applications developed within and for the academy, and finding these solutions should be a high priority for the academy, its libraries and publishers, to address.

Research Methods

The fourth strategic area that I would highlight is the advance of new discipline-based research methods. The development of search technologies will drive the scholarly use of massively digitized resources, but scholarly use will also shape and guide the development of particular technologies and applications for specific disciplinary pursuits. Disciplines will need to develop new and specialized methodologies—an informatics of standards and practices—to identify, mark up, and

explore the large volumes of digital information with which they each need to work: economists with tabular data in government publications; literature scholars with literary texts from various genres; social historians with contemporary accounts of various aspects of social life; ethicists with case studies of ethical dilemmas; art historians with evidence about the context of artists and their art; and so on. As scholars in various fields of study develop experience with these materials, the disciplines and subdisciplines will need to develop and codify practice.²⁸

New Publication Emphases

Fifth, as scholars begin to formulate how the use of databases and newly digitized materials could advance knowledge in their fields and begin to set discipline-based standards for how these materials should be organized for systematic, machine use, then we will likely need to pave the way for three further types of intensive scholarly activity that will increasingly dominate scholarly publishing:

- Editorial activity will shift, field-by-field, to the markup and online annotation of digital (or digitized) source materials to shape them for scholarly activity and machine processability in particular disciplines.
- Given appropriately edited and marked up resources, and proficiency in new methodological techniques, scholars will begin to generate and report results based on research using these methods. These reports will refer systematically to digitized sources and may incorporate them in various ways. They will make increasing use of the power of the computer to illustrate and represent ideas graphically; to simulate physical, biological, and social systems; to engage the reader interactively; and to document ideas encyclopedically with data and other evidence that are portable and recombinant in ways that allow arguments to be tested, proved, and extended. Complex works with these features will be the natural descendant of the monograph and the journal article, but will fit naturally in neither category.
- And, as scholars learn new ways of interpreting evidence and the scholarly record, they will be learning new ways to write and will need tools and processes to assist them and to make

dissemination throughout the academy easier and affordable in discipline-appropriate ways.

Researchers at the Institute for Advanced Technology in the Humanities at the University of Virginia and elsewhere, such as Robert Darnton, Edward Ayers, William Thomas, and others, have been modeling these new forms of scholarly practice in the humanities,²⁹ and there are already countless examples in the sciences.

New Collection Emphases

The sixth area that I would highlight as strategically crucial for the future of scholarly communications is the need for dramatic shifts in the emphases in collection building in libraries. If large quantities of published materials are available online through some common

interface, it will be increasingly hard to distinguish libraries based on their holdings of these materials. Instead, libraries and their institutions will increasingly be distinguished by the special collections of rare and unique materials that they hold and by the scholarly services they provide for these materials,

...libraries and their institutions will increasingly be distinguished by the special collections of rare and unique materials that they hold and by the scholarly services they provide for these materials, especially in conjunction with similar collections at museums and archives locally and around the world.

especially in conjunction with similar collections at museums and archives locally and around the world. Special collections are often inaccessible or under-processed, and the forms of description do not integrate well with other kinds of catalogs. Several institutions, including the Historical Society of Pennsylvania, the Virginia Historical Society, Columbia University, University of Virginia, and others, have been working together in recent years to develop innovative methods of appraising special collections for processing; others to simplify the cataloging.³⁰

Perhaps even more important is the need for more aggressive development of collections in new media. Recent and contemporary culture both here and abroad is documented in audio recordings, in still and moving images, broadcast media, and in various exclusively digital formats, such as large-scale, machine-generated scientific data sets, geographic information systems, simulations, Web pages, and weblogs. Scholars will increasingly need access to these materials for teaching and research.³¹ Concerted action is especially needed among libraries to ensure that these materials are actively and comprehensively collected and processed for scholarly use. Economies of scale, and the complexities associated with intellectual property rights management may prove that individual libraries need more centralized, collaborative mechanisms to achieve these objectives.

Interaction between Digital Library and Learning Management Systems

The last strategic issue that I would highlight is the need for more seamless interaction between digital library, publishing, and learning management systems. There is a pedagogical trend to incorporate the use of primary sources and research methods more deeply in the curriculum of higher education, and this trend will likely continue, but will also vary by discipline. As scholars in different fields gain experience with and develop discipline-based methodologies for using large-scale digitized content, as well as special collections and new media collections, they will need to incorporate the material and train students in the research methods. Demand will grow for deepening connections between digital library systems used for managing digital assets in various forms and combinations of licensed, digitized, and open access materials and learning management systems such as Sakai. Conversely, at least some of the content specifically created for teaching and learning will need to flow to digital library systems for long-term management and preservation. Essential for the effective management of the flows of content among digital library systems and between digital library systems are mechanisms, like Shibboleth, for building and expressing levels of trust between owners and users of the digital assets.³²

Conclusion

There is a view that the promise—or curse—of commercial digitization activities is that they will make libraries and publishers within the academy largely irrelevant. I hope you can tell from the strategic issues that I have highlighted for you this morning that I find such a view to be spectacularly uninformed and shortsighted. Rather, the promise—or the curse—is that scholarly communications has become a vastly more interesting enterprise than it has ever been. It is increasingly possible for scholars to have unprecedented access to the resources they need to engage issues that have remained elusive or even unthinkable. The challenge for us is to be both extraordinarily innovative and conservative at the same time: Innovative in that we must organize ourselves to take absolute best advantage of the opportunities; and conservative in that we must protect our gains and not screw up the scholarly process.

What about roles and division of labor? I cannot be prescriptive about what libraries or publishers should be or do in the new ecology of scholarly communications. There is just too much to do, and the competition from

higher education institutions abroad, especially in Asia is growing rapidly. The need is great for imagination, expertise, and other resources to be applied wherever they can be found, but please allow me to emphasize the powerful paradox of digital technologies.

Their use can be hugely liberating at the individual level, opening new realms of investigation for the scholar and new levels of educational attainment for the learner, especially those of lesser privilege here in the US as well as abroad. It is this spirit of democratization that is a treasure to see in evidence in the open access debates. At the same time, for reasons of economy and scale, the academy can unleash these democratizing activities only if it is able to consolidate core pieces of the infrastructure—the digitization process in the Google and OCA projects; software development for DSpace, Fedora, Sakai, and other open-source initiatives; and the aggregation of content in JSTOR, ARTstor

...for reasons of economy and scale, the academy can unleash these democratizing activities only if it is able to consolidate core pieces of the infrastructure....

and other databases. To build this common infrastructure, the need is huge for collaboration and collective organization involving shared financing and responsive governance at levels that are probably unprecedented, and this need raises another, larger question about how the academy can reorganize itself to accommodate efficiently and responsibly within its embrace entities that essentially outsource library, publishing and related functions that once were held closely within individual institutions. The California Digital Library is one model of outsourcing within a state system. JSTOR, ARTstor, and Portico, and LOCKSS represent yet other models, and the Mellon and Hewlett Foundations are experimenting with yet another in their jointly funded creation called Ithaka, which is designed to stitch together with common services ARTstor, JSTOR, Portico, and a family of other scholarly support entities.³³ These resources simply cannot take shape if they are imagined to be “one off,” or ad hoc organizations. Presidents, provosts, deans, scholars, librarians, press directors, editors, and technologists together must find ways within the larger academic community for their institutions to work together to realize the extraordinary economies of scale that are possible, and foundations like Mellon should not be seen as the “deep pockets” to which they turn to cover the costs of these entities, but as catalysts in the necessary effort to establish them financially and organizationally as new modes of ongoing operation in higher education.

Let me now leave you with a cautionary tale from Adrian John's *Nature of the Book*. In 19th-century England, there arose a group called the Society for the Diffusion of Useful Knowledge. Worried that an educated working

class could be a dangerous force in society, it resolved to swamp the country with cheap magazines—the Penny Cyclopaedia and the Penny Magazine—that contained absolutely nothing “to excite the passions.” To achieve this mission, the society was the first group to make full, industrial use of the steam press, a remarkably cost-effective technology at large scale. By 1832, the society’s magazine was “by far the most extensively circulated periodical works that issue from the press.” It estimated its readership at the then unprecedented figure of one million.

However, for all the attention to cost-effectiveness, critics of all persuasion attacked the project and it eventually failed. Conservatives were convinced that the project dispersed unnecessary ideas that might still prove dangerous. Radicals, on the other hand, complained that the magazines contained no really useful knowledge. Instead, they said, rather than meeting demand, the society sought nothing more than to “stuff our mouths with Kangaroos.”³⁴

As discussions continue about how most effectively to manage scholarly digital assets, let us not fall into the trap of the Society for the Diffusion of Useful Knowledge and lose sight of the ultimate objective: meeting demand for useful knowledge. Let us be on the lookout for the “Kangaroos.”

—Copyright 2006 Donald J. Waters

The above remarks were presented at “Managing Digital Assets: Strategic Issues for Research Libraries,” Washington, DC, October 28, 2005. This forum was sponsored by ARL, CNI, the Council on Library and Information Resources, and the Digital Library Federation. The proceedings from the forum are available online at <http://www.arl.org/forum05/>.

¹ Adrian Johns, *The Nature of the Book: Print and Knowledge in the Making* (Chicago: University of Chicago Press, 1998): 32–33.

² Roger Noll and W. Edward Steinmueller, “An Economic Analysis of Scientific Journal Prices: Preliminary Results,” *Serials Review* 18, no. 2 (1992): 32–37.

³ President’s Science Advisory Committee, *Science, Government, and Information: The Responsibilities of the Technical Community and the Government in the Transfer of Information* (Washington, DC: The White House, 1963); National Enquiry, *Scholarly Communication: The Report of the National Enquiry* (Baltimore: Johns Hopkins University Press, 1979); see also Anthony Cummings et al., *University Libraries and Scholarly Communications* (Washington, DC: Association of Research Libraries, 1992); and Richard Ekman and Richard E. Quandt, eds., *Technology and Scholarly Communication* (Berkeley: University of California Press, 1999).

⁴ Lee C. Van Orsdel and Kathleen Born, “Choosing Sides—Periodical Price Survey 2005,” *Library Journal* (April 15, 2005), <http://www.libraryjournal.com/article/CA516819.html>.

⁵ See, generally, Charles W. Bailey, Jr., *Open Access Bibliography: Liberating Scholarly Literature with E-Prints and Open Access Journals* (Washington, DC: Association of Research Libraries, 2005), <http://www.escholarlypub.com/oab/oab.pdf>; and Declan Butler and Philip Campbell, “Future E-Access to the Primary Literature,”

Nature Web Debates (April 5, 2001), <http://www.nature.com/nature/debates/e-access/introduction.html>; Declan Butler, “Introduction,” *Nature Web Focus* (March 19, 2004), <http://www.nature.com/nature/focus/accessdebate/1.html>; and John Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship* (Cambridge: MIT Press, 2006).

⁶ James L. Morrison and Peter Suber, “The Free Online Scholarship Movement: An Interview with Peter Suber,” *The Technology Source* (September/October 2002), http://technologysource.org/article/free_online_scholarship_movement.

⁷ David Malakoff, “Scientific Societies Lay Out ‘Free Access’ Principles,” *Science Magazine* 303 (March 26, 2004): 1959.

⁸ See, for example, Stevan Harnad, “How to Fast-Forward Learned Serials to the Inevitable and the Optimal for Scholars and Scientists,” *The Serials Librarian* 30, no. 3/4 (1997): 73–81, http://cogprints.ecs.soton.ac.uk/archive/00001695/00/harnad97.1_earned.serials.html.

⁹ See Paul Ginsparg, “Winners and Losers in the Global Research Village,” *The Serials Librarian* 30, no. 3/4 (1997): 83–95, <http://arxiv.org/blurb/pg96unesco.html>; Paul Ginsparg, “Creating a Global Knowledge Network,” *BMC News and Views* 1, no. 9 (Dec. 12, 2000), <http://www.biomedcentral.com/1471-8219/1/9>; and Herbert Van de Sompel et al., “Rethinking Scholarly Communication: Building the System that Scholars Deserve,” *D-Lib Magazine* 10, no. 9 (2004), <http://www.dlib.org/dlib/september04/vandesompel/09vandesompel.html#1>. Also see Joseph J. Esposito, “The Devil You Don’t Know: The Unexpected Future of Open Access Publishing,” *First Monday* 9, no. 8 (2004), http://firstmonday.org/issues/issue9_8/esposito/index.html; and Jean-Claude Guédon, “The ‘Green’ and ‘Gold’ Roads to Open Access: The Case for Mixing and Matching,” *Serials Review* 30, no. 4 (2004): 315–328.

¹⁰ Clifford Lynch, “Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age,” *ARL Bimonthly Report*, no. 226 (February 2003), <http://www.arl.org/newsltr/226/ir.html>; Steve Hitchcock, “Metalist of Open Access E-Print Archives: The Genesis of Institutional Archives and Independent Services,” *ARL: A Bimonthly Report*, no. 227 (April 2003), <http://www.arl.org/newsltr/227/metalist.html>; Clifford A. Lynch and Joan K. Lippincott, “Institutional Repository Deployment in the United States as of Early 2005,” *D-Lib Magazine* 11, no. 9 (2005), <http://www.dlib.org/dlib/september05/lynch/09lynch.html>.

¹¹ Terry Ehling, “The Development of an Open Source Publishing System at Cornell and Penn State Universities,” *ARL Bimonthly Report*, no. 237 (December 2004), <http://www.arl.org/newsltr/237/opensource.html>; Paul Resnick, “Implications of Emerging Recommender and Reputation Systems,” in *Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications: Proceedings of a Symposium* (Washington, DC: National Academies Press, 2004): 68–70, http://books.nap.edu/html/e_journals/ch6.html; Richard Luce, “Preprint Servers and Extensions to Other Fields,” in *Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications: Proceedings of a Symposium* (Washington, DC: National Academies Press, 2004): 70–72, http://books.nap.edu/html/e_journals/ch6.html.

¹² Jason Epstein, *Book Business: Publishing Past, Present, and Future* (New York: W.W. Norton, 2001): 173–174.

¹³ Roger Schonfeld et al., *The Nonsubscription Side of Periodicals: Changes in Library Operations and Costs between Print and Electronic Formats* (Washington, DC: Council on Library and Information Resources, 2004), <http://www.clir.org/pubs/reports/pub127/pub127.pdf>.

¹⁴ Michael A. Keller et al., “What Is a Library Anymore, Anyway?” *First Monday* 8, no. 5 (2003), http://www.firstmonday.org/issues/issue8_5/keller/index.html.

¹⁵ See Donald J. Waters, ed., “Urgent Action Needed to Preserve Scholarly Electronic Journals,” notes of a meeting of university librarians and academic administrators held at the offices of The Andrew W. Mellon Foundation on September 13, 2005, <http://www.diglib.org/pubs/waters051015.htm>.

CURRENT ISSUES

Continued

- 16 Ed Pentz, "CrossRef: A Collaborative Linking Network," *Issues in Science and Technology Librarianship* (Winter 2001), <http://www.istl.org/01-winter/article1.html>; Ed Pentz, "Evolution and Revolution: Pragmatism versus Dogmatism," *Nature Web Debates* (April 5, 2001), <http://www.nature.com/nature/debates/e-access/Articles/pentz.html>.
- 17 Vicky Reich, "Lots of Copies Keep Stuff Safe as a Cooperative Archiving Solution for E-Journals," *Issues in Science and Technology Librarianship* (Fall 2002), <http://www.istl.org/02-fall/article1.html>; Eileen Fenton, "Preserving Electronic Scholarly Resources: An Overview of Portico," presented at Coalition for Networked Information Fall Task Force Meeting, 2005, <http://www.cni.org/tfms/2005b.fall/abstracts/PB-preserving-fenton.html>; see also Johan F. Steenbakkers, "Digital Archiving in the 21st Century: Practice at the National Library of The Netherlands," *Library Trends* 54, no. 1 (Summer 2005): 33–56.
- 18 William LeFurgy, "Building Preservation Partnerships: The Library of Congress National Digital Information Infrastructure and Preservation Program," *Library Trends* 54, no. 1 (2005): 163–172, http://www.ndiipp.uiuc.edu/pdfs/library_trends_ndiipp.pdf.
- 19 See, for example, Stephen Pinfield and Hamish James, "The Digital Preservation of e-Prints," *D-Lib Magazine* 9, no. 9 (2003), <http://www.dlib.org/dlib/september03/pinfield/09pinfield.html>.
- 20 Lorcan Dempsey, "Libraries in the Age of Amazoogole: Some Issues and Responses: The Example of OpenWorldCat," presented at the JISC/CNI Conference, June 8–9, 2004, Brighton, England, <http://www.oclc.org/research/presentations/dempsey/brightonowc.ppt>.
- 21 David Kesmodel and Vauhini Vara, "Building an Online Library, One Volume at a Time," *Wall Street Journal* (November 9, 2005), http://online.wsj.com/public/article/SB113111987803688478-VNpw62xi_JA4avE8cxOZf0pf_nM_20061109.html?mod=blogs.
- 22 Siva Vaidhyanathan, "A Risky Gamble with Google," *Chronicle Review* 52, no. 15 (2005): B7.
- 23 Joseph Esposito, "The Processed Book," *First Monday* 8, no. 3 (2003), http://www.firstmonday.org/issues/issue8_3/esposito/index.html.
- 24 Eric Schmidt, "Books of Revelation," *Wall Street Journal* (October 18, 2005): A18, <http://googleblog.blogspot.com/2005/10/point-of-google-print.html>.
- 25 See, for example, David Lindahl and Jeff Suszczynski, "Rethinking Metasearch for a Better User Experience," presented at conference on "OpenURL and Metasearch: New Standards, Current Innovations, and Future Directions," National Information Standards Organization, Washington, DC, September 20, 2005, http://www.niso.org/news/events_workshops/OpenURL-05-ppts/2-5-lindahl-susz.ppt.
- 26 Lorcan Dempsey, "The Library and the Network: Flattening the Library and Turning It Inside Out," presented at the "Access 2005" conference, Edmonton, Alberta, Canada, October 19, 2005, <http://www.oclc.org/research/presentations/dempsey/edmonton.ppt>.
- 27 See, for example, Jamie Callan and Alan Smeaton, eds., *Joint NSF-EU DELOS Working Report: Personalization and Recommender Systems in Digital Libraries* (Sophia Antipolis, France: European Research Consortium for Informatics and Mathematics (ERCIM), 2003), <http://www.ercim.org/publication/ws-proceedings/Delos-NSF/Personalisation.pdf>.
- 28 See, for example, Jerome McGann, "Culture and Technology: The Way We Live Now, What Is to Be Done?" *Interdisciplinary Science Reviews* 30, no. 2 (2005): 179–189, <http://www.ingentaconnect.com/content/maney/isr/2005/00000030/00000002/art00008>; and more generally for the humanities, Susan Schreibman et al., eds., *A Companion to Digital Humanities* (Oxford: Blackwell, 2004).
- 29 Robert Darnton, "Presidential Address: An Early Information Society: News and the Media in Eighteenth-Century Paris," *American Historical Review* 105, no. 1 (2000), <http://www.historycooperative.org/journals/ahr/105.1/ah000001.html>; William G. Thomas, III and Edward L. Ayers, "The Differences Slavery Made: A Close Analysis of Two American Communities," *American Historical Review* 108, no. 5 (2003), <http://www.historycooperative.org/ahr/elec-projects.html>.
- 30 Rachel Onuf, "Archival Collection Management Systems: A Model for Prioritizing Backlogs," presented at the International Congress on Archives, Vienna, Austria, August 25, 2004, http://www.wien2004.ica.org/imagesUpload/pres_206_ONUF_ZUSABERK01.pdf; Bradley D. Westbrook et al., "Archivists' Toolkit: Toward Streamlined Archival Workflow," presented at the annual meeting of the Society of American Archivists, August 20, 2005, http://archiviststoolkit.org/AT_SAA_2005/AT_presentation_SAA2005.pdf.
- 31 Roy Rosenzweig, "Scarcity or Abundance? Preserving the Past in a Digital Era," *American Historical Review* 108, no. 3 (2003), <http://www.historycooperative.org/journals/ahr/108.3/rosenzweig.html>.
- 32 See, for example, Neil McLean and Clifford Lynch, "Interoperability between Information and Learning Environments—Bridging the Gaps: A Joint White Paper on behalf of the IMS Global Learning Consortium and the Coalition for Networked Information" (Dublin, Ohio: Online Computer Library Center, 2004), http://www.imsglobal.org/digitalrepositories/CNIandIMS_2004.pdf; and Dale Flecker and Neil McLean, eds., *Digital Library Content and Course Management Systems: Issues of Interoperation: Report of a Study Group* (Washington, DC: Digital Library Federation, 2004), <http://www.diglib.org/pubs/cmsdl0407/index.htm>.
- 33 William G. Bowen, "New Times Always; Old Times We Cannot Keep," *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, no. 243 (December 2005): 1–16, <http://www.arl.org/newsltr/243/ARLBR243.pdf>.
- 34 Johns: 629–630.