Preserving the Scholarly Commons in the Digital Era

by

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

Preliminary Draft

Not for quotation or citation
without permission of the author

Workshop on Scholarly Communications as a Commons
Indiana University
Bloomington, Indiana
March 30-April 2, 2004

Preserving the Knowledge Commons: The Scholarly Record in the Digital Era

Not long ago, Anthony Grafton, the distinguished Princeton historian, published a history of the footnote. An intellectual tool that is "the humanist's rough equivalent of the scientist's report on data," the footnote offers "the empirical support for stories told and arguments presented." No doubt we all remember our own experiences of awe and wonder when we learned how to interpret a footnote and so began to understand the mechanics of scholarly reference. However, according to Grafton, "no one has described the way that footnotes educate better than Harry Belafonte, who recently told the story of his early reading of W. E. B. DuBois."

As a young West Indian sailor, Belafonte learned to read critically when he figured out how the footnote opened a world of learning. "I discovered," Belafonte said, "that at the end of some sentences there was a number and if you looked at the foot of the page the reference was to what it was all about—what source DuBois gleaned his information from." However, Belafonte did not find the task of learning from references to be easy at first and was stymied by the methods that DuBois used to cite his references. Trying to track them down, he says that he went to a library in Chicago with a long list of books. "The librarian said, 'that's too many, young man. You're going to have to cut it down.' I said, 'I can make it very easy. Just give me everything you got by Ibid.' She said, 'There's no such writer.' I called her a racist. I said, 'Are you trying to keep me in darkness?' And I walked out of there angry.".

Of course, footnotes are not the only or, in a variety of research and educational contexts, even the best method of reference. Moreover, as the Belafonte story indicates, there can be many obstacles in tracing down a reference path. However, as Grafton concludes in his study, the footnote is a critical part of the scholarly apparatus because it is such a clear and efficient mechanism to link one piece of scholarship with what its author has identified as the key reference points for the work. It serves as a guarantee, Grafton says, "that statements about the past derive from identifiable sources. And that is the only ground we have to trust [those statements]" (Grafton 1997: vii, 233-235).

In other words, when scholars use systems of reference to link one work to another, they establish and exercise underlying fabrics of trust. These fabrics serve to tie researchers to other researchers, teachers to students, and creators to users over time and place into durable and productive scholarly communities. The linked works represent the common pools of knowledge—the scholarly commons—over which members of these communities labor to produce new knowledge. And the links work, the trust endures, and the commons nourishes the intellectual life if and only if

cited material is preserved so that, when a link is made, the reader is able to check a reference at the other end at will.

Academic libraries have traditionally preserved the scholarly record in printed form by buying books and journals from publishers for their local researchers, teachers, and students. They store these works in protective environments, fix bindings and pages when necessary, and microfilm or digitize those volumes in danger of deterioration. Today, increasing numbers of scholars are contributing articles to electronic journals and taking part in projects to publish electronic books, and they support their scholarship with citations to related digital materials as well as to more traditional sources. Such electronic scholarship is as important for the cultural record and the building of knowledge as printed publications have been, and is therefore as important to preserve. But libraries generally do not buy electronic journals and books. They rent them. So who is taking responsibility for preserving these materials?

Although it is relatively easy to make a persuasive case for why digital preservation is necessary, an impressive array of factors and incentives—including the fundamental shift from buying to renting—leads otherwise well-intentioned actors in different directions. Meanwhile, digital materials are proving to be fragile and fleeting with potentially serious consequences for the scholarly commons. Brewster Kahle, who founded the Internet Archive to preserve portions of the Web, estimates that a Web object now has an average life expectancy of 100 days (Weiss 2003). Mortality is also high for Web-based scholarly literature. A study published in *Science* in October 2003 found that more than 30 percent of the articles in selected high impact medical and scientific journals contained one or more Internet references, but "the percentage of inactive Internet references increased from 3.8% at 3 months to 10% at 15 months and to 13% at 27 months after publication" (Dellavalle 2003:787). A similar study conducted in 2001 found that the percentage of inactive Internet references increased from 23 percent at two years to 53 percent at seven years after publication (Lawrence 2001). With additional effort, many of the works cited in the inactive references could still be found, but the results of these studies clearly indicate that the digital ecology of the scholarly commons is less than stable, and its preservation is far from assured.

In this paper, I focus specifically on the problem of preserving electronic scholarly journals (e-journals). To provide a framework for analyzing the problem and possible solutions, I first define it as a problem of preserving a commons, and then explore key roles and organizational models in the preservation process. I conclude by identifying key features of what might emerge as community-based preservation efforts.

E-journal preservation as a commons problem

In the fall of 2000, The Andrew W. Mellon Foundation invited seven of the nation's leading universities, along with publishers that they each selected, to participate in a preservation planning process (Cantara 2003). Together, the participants would develop and share detailed understandings of the requirements for setting up and implementing trustworthy archives for the preservation of electronic journals, create technology to facilitate the archiving process; and organize the implementation and operation of electronic journal archives. Although they demonstrated in many ways the technical feasibility of preserving electronic journals, most of these seven planning projects stalled when they ran smack into the some of the classic problems of the political economy of public goods: What are the incentives for individuals and institutions to participate in the provision and maintenance of a good when others cannot be readily excluded from enjoying the benefit? What are the organizational options? What are sustainable funding plans?

Commons—or more specifically common pool resources—are a kind of modified public good. They share with public goods the feature that it is difficult to exclude beneficiaries, but differ in that use may reduce the availability of the resource to others (Ostrom, *et al.* 1999: 278). Knowledge in the abstract, such as the theory of relatively, is strictly speaking a public good, because it is difficult to exclude people from benefiting from the theory and use of the theory does not diminish its availability to others. Knowledge in the form of specific works, such as articles in electronic journals, resembles a public good because it is also difficult to exclude beneficiaries who can readily copy, discuss or otherwise disseminate the material. Copyright protection is meant to provide incentives to those who might be deterred by the threat of copying from contributing in the form of publications to the common pool of knowledge. However, once a scholarly work is available in the form of a published electronic artifact, it can, like other kinds of common pool resources, be used up and, as linked references in e-journals, may simply disappear. To have its beneficial effects, a published work needs to be available to the broadest possible audience, but access is not equivalent to preservation and at the limit, unless care is taken, free or open access of a common pool resource may actually hamper preservation. Insuring against the loss of electronically published works is a common-pool resource problem that requires special attention.

To understand the nature of the problem, let us examine the idea that the preservation, or "archiving," of electronic journals and other forms of electronic publications is in fact insurance against loss. Is preservation really like insurance, in the sense of fire or life insurance? Would a business approach based on an insurance model induce people to take on responsibility for archiving? If you have fire insurance and your house burns down, you are protected. If you have life insurance and you die, your heirs benefit. There is an economy in these kinds of insurance that induces you to buy. If you fail to buy, you are simply out of luck; you are excluded

from the benefits. Unfortunately, the insurance model for preserving electronic journals is imperfect, because insurance against the loss of information does not enforce the exclusion principle.

A special property of archiving is that if one invests in preserving a body of electronic journals and the works are eventually lost to others who did not take out the insurance policy, the others are not excluded from the benefits, because the knowledge in the works still survives. Because free riding is so easy, there is little economic incentive to take on the problem of digital preservation. Potential investors conclude: "it would be better for me if someone else paid to solve the archiving problem." As we have seen, one of the defining features of a common pool resource is that it is difficult and costly to exclude beneficiaries.

Given the huge free-riding problem associated with the maintenance of the knowledge commons, what are the alternatives? Reflecting in part on this problem, Garrett Hardin despaired of solutions. "Ruin," he wrote in "The Tragedy of the Commons," "is the destination toward which all men rush, each pursuing his own interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all" (1968: 1244). Hardin echoed Thomas Hobbes, who lamented the state of nature, a commons in which people pursue their own self-interest and lead lives that are "solitary, poore, nasty, brutish, and short" (1651: 65). Focused on preserving digital information in 1996, the Task Force on Archiving of Digital Information echoed both Hobbes and Hardin in writing that "rapid changes in the means of recording information, in formats for storage, in operating systems, and in application technologies threaten to make the life of information in the digital age 'nasty, brutish, and short'"(Waters and Garrett 1996, 2).

One of Hardin's solutions to the tragedy of the commons was, like Hobbes's, to rely on the leviathan—the coercive power of the government. Governments, in fact, have funded many of the early efforts to create digital archives (Beagrie 2003; Library of Congress 2002). Hardin's other solution was to encourage privatization, trusting in the power of the market to optimize behavior and preserve the commons. Efforts such as Brewster Kahle's Internet Archive demonstrate the kinds of contributions that private investment could make.

Certainly, both the government and private interests have roles to play in preserving the scholarly commons, but substantial experimental and field research in the political economy of public goods has also shown Hardin's pessimism about the prospects of maintaining common pool resources goods to be unwarranted. Case after case demonstrates that groups of people with a common interest in a shared resource will devise and agree upon community-based mechanisms for controlling and financing the preservation of the resource (Ostrom *et al* 2002; Dietz 2003). However, understanding the potential interaction of government, private, and

community interests in the systematic preservation of a digital scholarly commons requires a close analysis of potential roles, responsibilities and models of organization.

Preservation roles, responsibilities and models of organization

According to Brian Lavoie (2003), there are essentially three roles at play in the archiving equation. Lavoie uses slightly different labels, but I would refer to them as Producer, Consumer, and Archive. The Producer is the individual or set of individuals who generates an information object and is initially responsible for the bundle of ownership rights associated with the object. The Consumer is the individual or set of individuals that comprises the Public or Publics interested in the long-term preservation of an object. I use the word "consumer" deliberately to indicate the potentially complex relationship in which the producer may be selling, licensing, or otherwise supplying services to the Consumer based on the very same object that the Consumer wants to be preserved. And, as I would define it, the Archive is responsible for exercising the rights and duties of preserving the cultural, historical, or scholarly record.

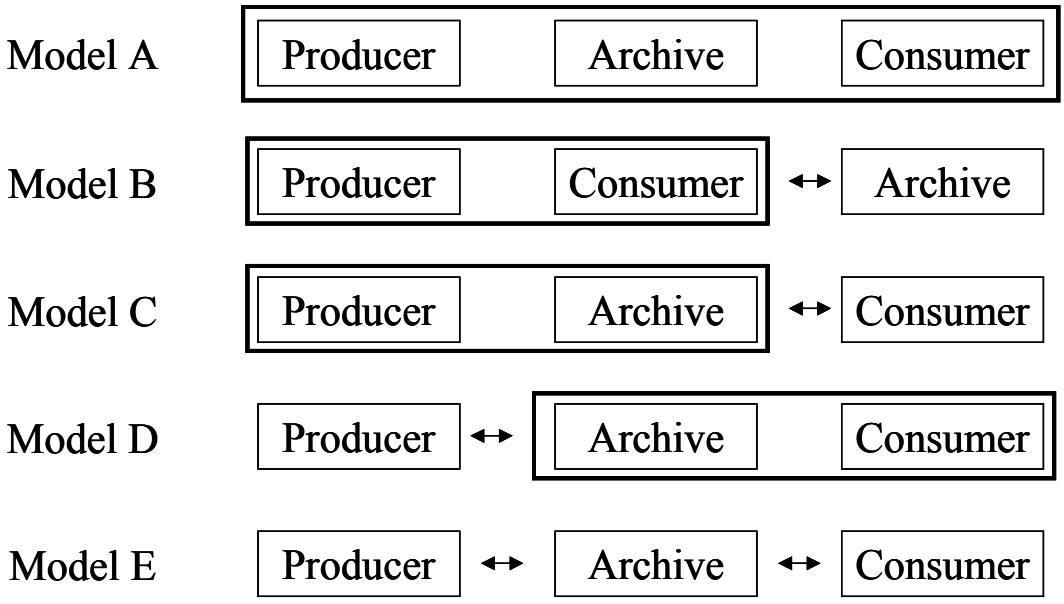

Figure 1. Organizational Models

As Lavoie observes, these three roles could logically be combined in five different ways, representing distinct organizational models (see Figure 1). The real world, of course, is a lot messier than these simple representations suggest, but there

is a heuristic value in considering these abstractions because they help us identify some of the key issues. I am departing from Mr. Lavoie's analysis here to suggest that two of the models, which I have labeled Models A and B represent forms of institutional archives.

Institutional archives

The key defining quality of both models is that the producer of the information objects and the consumer of the preservation service belong to the same institution. The institution in effect has a compelling interest and incentive to preserve the objects that it produces. The difference between the two models is that in the one case—Model A—the archive is housed within the boundaries of the institution, while in the other case—Model B—the archive is outsourced to some third party provider.

Both of these cases are relatively unproblematic. Because the institution controls its own finances and organization, it controls the demand for archiving, the allocation of roles and responsibilities, and the wherewithal to enable actors within the organization to carry out their responsibilities. Note, however, that if the institution is a complex one, and we take a perspective from within the organization, it may well be that to many of the internal actors the model would appear indistinguishable from Model 5 with many of the difficult problems that it entails.

Note also that one of the heuristic values of modeling roles and responsibilities in this schematic way is that it allows us to distinguish at least two senses in which institutional archives are used, often ambiguously, in current discourse. On the one hand, they refer in a strict sense to the case of an institution managing its own records. The institution is its own customer for purposes of archiving, and is not concerned with a broader public. Much of the early implementation of DSPACE as an institutional repository was designed solely to address the internal needs of MIT, with departments and groups within the institution contracting with the Library to archive as an internal record of digital products that they have generated (Barton and Walker 2003).

On the other hand, a good deal of the rhetoric about institutional repositories, particularly in the recent position paper produced by SPARC, suggests that institutional archives could do much more, including holding copies of all the published papers produced by their faculty, and thereby appealing to a demand for preservation from a much larger customer base that extends well beyond the bounds of the institution (Crow 2002). There is little evidence that such a vision is feasible in the short term, but I don't mean to suggest that the vision is not worth pursuing. Rather, in terms of the formal models outlined here, I would merely observe that the SPARC vision is not strictly speaking an institutional archive; instead, when

universities embrace such a vision, the relevant actors will likely share roles and responsibilities more like those in Model C, which I would call producer archives.

Producer archives

Model C represents those cases in which the producer and the archive are aligned institutionally to preserve a portion of the cultural record for a broad consumer base. Besides the SPARC vision of colleges and universities creating archives of the publications that their faculty authors produce, other examples of producer archives would be publisher archives and so called author self-archives. Is preservation in the mission of such producers and are they credible archives?

Universities as the producers of knowledge have traditionally relegated collecting and preserving the scholarly literature to their libraries. Libraries, in turn, have taken it as their mission to embrace collections that are broadly useful as resources for research and teaching within the institution, rather than to focus on archiving the published output of their faculty. Shifting the preservation mission of academic institutions is not inconceivable but, as Clifford Lynch (2003) has pointed out, would likely require significant, and potentially costly, cultural, policy, and technical changes that could distract from the larger academic mission of encouraging innovation and the expansion of knowledge, and may require federating technologies that either do not exist or are currently too immature to be useful.

For many publishers, other than some large scholarly societies, it is doubtful whether preservation is any more than what economists would call a "positive externality," something that would be nice to achieve if someone else paid for it, but not worth any significant expenditures of their own. However, publishers are certainly not indifferent to the issue of archiving. For as long as their databases are commercially viable, publishers have a strong interest in preserving the content—either themselves or through a third party. Scholarly publishers also have an incentive to contribute to preservation activity in the interests of their authors, who want their works to endure, be cited, and serve as building blocks for knowledge. Still, a primary concern about the viability of producer archives is whether the material is in a preservable format and can endure outside the cocoon of the publisher's proprietary system. One necessary ingredient in a proof of archivability is the transfer of data out of their native home into external archives, and as long as publishers refuse or are unable to make such transfers, this proof cannot be made.

Another concern about producer archives is more subtle, and perhaps more pernicious in its implications for the future of the scholarly commons in electronic form. In part, because electronic publications are generally maintained online, rather than being physically transferred as paper publications are in a sale, publishers appear to be more vulnerable to legal demands, editorial second-guessing, and other activities

that result in the removal of materials from the publishers archive. Last year, in a thread called "the vanishing act" correspondents on the LibLicense listserv documented multiple instances in which Elsevier had succumbed to pressures to remove articles. In relation to the overall size of Elsevier's database, the number of vanishing articles was, of course, relatively small, and it is doubtful that Elsevier is the only publisher that has been subject to such pressures, but the consequence of removal is that it produces a "Swiss cheese" effect in the scholarly record and casts doubt on the ability of publishers in general to preserve the integrity of the commons, at least on their own.

The LibLicense discussion about publisher-removed articles then prompted James O'Donnell, the provost of Georgetown University to observe that the "vanishing act" discussion "is disturbing, because it is the tip of the iceberg, I think: If for fairly transient reasons, publishers will pull articles, when might not publishers prove unreliable for other reasons?" He went on then to highlight how the failure to account for reliable preservation is one of the most poorly examined open spaces under the head of steam known as "author self-archiving." O'Donnell writes:

> But the question that follows on this discussion for me is this: If we were to ask that not publishers but authors be the guarantors of permanence, self-publishing or publishing in institutional repositories where the author retains control over the copyright and disposition of his/her material—what protection do we then have to assure us that articles will remain archived, unchanged, in perpetuity? Are there articles I have written that I wouldn't mind disappearing? Actually, yes. Are there pieces of articles that I would quietly change if I could? Well, interesting thought, sure.

Consumer archives

Let me now turn briefly to Model D, which represents what I would suggest is a consumer archive. In the digital realm, as with other forms of information, the passions and interests of what Edward Tenner has called "freelance selectors and preservers" will almost surely result in valuable collections of record (2002, 66). Just as publishers undoubtedly have a role in digital archiving, so too will individual consumers. However, just as there is reason to question the commitment of producers to the long-term task of preservation, so too consumer archives are subject to similar, and perhaps even greater, concern, and provision must be made to ensure the eventual transfer of archived materials to archives capable of providing long-term care.

Community-based archives

This brings me to the last, and perhaps most interesting and complex organizational model, Model E. In this case, each of the three significant roles are played by independent actors. Ideally, there will emerge a network of competent digital archives that would be responsible for preserving electronic journals and other digital materials of cultural and scholarly significance. Indeed, if the model being developed by the Library of Congress (2002) eventually succeeds, the archival function itself may depend on a complex and distributed division of labor among parties with various responsibilities for selection and custodianship, security, and repositories. But the key organizational feature of this model for the preservation of electronic journals is that members of the scholarly community, including producers, especially publishers, consumers as represented by scholars and their academic institutions, and libraries would find ways jointly to solve this pressing problem.

The Mellon Foundation also expects to play a supporting role as part of the community, especially given its long-standing philanthropic interest in the preservation of the cultural record as a condition of excellence in higher education. However, it is looking, as it does in nearly all cases of support, for ways to promote a self-sustaining, businesslike activity. It cannot in this, or in any other initiative, support long-term operating costs without compromising its mission. As a result, the Foundation seeks to foster the development of communities of mutual interest around preservation, help legitimize archiving solutions reached within these communities, and otherwise stimulate the necessary support from within the scholarly community. The premise of the Mellon e-journal planning projects, which I mentioned above, was that concern about the lack of solutions could be addressed only by hard-nosed discussions among stakeholders about what kinds of division of labor and rights allocations are practical, economical, and trustworthy, and from those planning projects two fledgling preservation services were born.

One was the development of E-Archive, a new organization that is affiliated with JSTOR and is being designed to preserve the source files used to publish electronic journals. Since its inception in late 2002, E-Archive has developed a business relationship with 10 publishers. It has developed mechanisms for transferring data from these publishers, and has designed and constructed a prototype repository. It has verified through a detailed study that a shift from print to electronic journals would generate huge savings in non-subscription processing and storage costs within libraries (Schonfeld 2004), and it is now negotiating with publishers and libraries to determine what services it can offer that will command the fees needed to sustain the archive.

The other initiative is based at Stanford University and is developing an archiving system called LOCKSS, for Lots Of Copies Keeps Stuff Safe. In the LOCKSS system, a low-cost Web crawler under the control of a participating library

is used for systematically capturing presentation files—Web-based materials that publishers use to present journal content to readers. Publishers allow the files to be copied and stored in Web caches that are widely distributed on local campuses but highly protected. The caches communicate with each other through a secure protocol, checking each other to see whether files are damaged or lost and repairing any damage that occurs. Caching institutions have the right to display requested files to those who are licensed to access them if the publisher's site is unavailable and to provide the local licensed community the ability to search the aggregated files collected in the institutional cache. Much work remains, but Stanford has attracted more than 80 libraries and more than 50 publishers to test the system, and expects LOCKSS to be preserving 100 electronic journal titles from eight to ten publishing platforms by spring 2004, when a full production system is released. Like E-Archive, however, LOCKSS has yet to generate the revenue needed from the community to sustain the enterprise.

Key Features for Preserving the Scholarly Commons

Scholarly communications is today in considerable flux. Repeated extensions of the term of copyright appear to be putting the scholarly and other information commons into private hands. Scholarly journal prices continue to spiral out of control, prompting a surge of innovation and experimentation with alternative publishing and distribution models. Meanwhile, scholarship is moving inexorably to a digital basis, with scholars increasingly producing and citing digital resources.

In all of this turmoil, there is almost no accounting, fiscal or otherwise, for the long-term preservation of scholarly resources, such as e-journals. It is that continuing neglect that may, in the end, pose the most serious threat to the scholarly commons. Given the analysis of roles and organizational models for the preservation of e-journals in the context of the scholarly commons, what features might be most usefully highlighted to chart a path forward?

Definition of Archives

The first feature I would suggest is that the role of e-journal archives must be narrowly defined in terms of the rights and duties needed to preserve the historical, cultural, or scholarly record. Others are better qualified than I to comment on whether or not the various recent revisions to the copyright law are constitutional or how well they balance private interest and the public good, but I am deeply concerned that activity aimed at preserving digital materials is not sufficiently distinguished in debates about intellectual property and so the implications of new law or court review for preservation receive scant attention and little protection in what are otherwise sweeping and potentially far-reaching changes. Part of the

problem is that, as a community, we have not been very rigorous in defining the archival role with respect to digital information.

Over the last decade, the semantics of the word "archives" has grown increasingly complex. The narrow, traditional definition of an archives as a repository with a long-term responsibility for preserving the cultural record has been extended in such uses as the "Open Archives Initiative," "scholar self-archives," and "computer archives," to refer simply to collections of interest or even more simply to ordinary daily backup systems. These loosely-defined senses are often used interchangeably or in association with the more rigorous definitions, and so tend to generate more confusion than clarity.

Here, for example, is a fairly common definition of the mission of a digital archive, which appeared in a recent report of a Mellon-funded project: "To ensure the long-term survival and broad availability of digital information." I will return shortly to the highly problematic assertion about access in the phrase "broad availability," but even the term "long term survival" is overly broad because everything cannot be saved and the archival function, in a strict definition, is specifically associated with the highly particular and selective function of identifying and preserving historical, cultural, or scholarly records. Preservation is a daunting task in any case. When the definition of archives is not restricted to this highly focused objective it is hard for policy-makers, such as members of congress, judges, and especially the clerks who do their research for them, to see the distinct value of the task and take it seriously enough to consider its implications when making decisions that may affect the scholarly community's ability to manage and preserve its cultural record.

Legal Protection

A second feature that I would highlight is that e-journal archives, strictly defined, may need legal protection from the negative effects of liability and other tort actions against publishers. If transfer of an e-journal from a producer to an archive is proof of archivability, then it behooves the archives to institute the transfer as soon after formal publication as possible to ensure against producer actions that might change or remove material. However, even if a hand-off were immediate, a license or other form of contract between the producer and the archive may govern the transfer, and may not protect the archive from requests to withdraw material in the same way that the sale and physical transfer of a printed publication would (Ayre and Muir 2004). In other words, the interactions among contract, tort actions, business decisions, and copyright may leave long-term archives exposed in the digital environment in ways that they are specifically protected in other media, at least in U.S. copyright law.

Given this definition of the problem, the Foundation recently commissioned a comparative study that is looking carefully at the legal interaction of contract, tort, and copyright in the US and a few other countries. If this study proves that there is a deeper structural problem, it may be necessary to create or employ appropriate legal and policy constructs, analogous to those that accompany the sale of a paper copy of a journal with an offending article, that would shield the archive from demands to change or withdraw material from online view. One possibility may be to begin to articulate "safe harbor" principles about intellectual property rights that could form the basis of digital archiving agreements among interested parties.

In building JSTOR and ARTstor as archival resources, the Foundation has found that content owners are much more comfortable with agreements that limit uses of intellectual property to not-for-profit educational purposes than they are with agreements that leave open the possibility of creating competing commercial profit-making access to the property. Lawrence Lessig has also recently argued for the utility of the distinction between not-for-profit educational uses and other kinds of uses of intellectual property (2001: 249-261). Because educational use is certainly consistent with the Constitutional mandate for intellectual property law in the United States to promote "the progress of science and useful arts," perhaps it is time to build a safe-harbor framework for digital archiving on just such a distinction.

Business model

The third and final feature that I would highlight is the need for an adequate and sustainable business model. In order for the scholarly community to afford archives for e-journals and other digital resources, key questions still need immediate and imaginative attention: What rights and privileges would archives need to be able to preserve published content in digital form? Can ways be found to apply the exclusion principle in such a manner that it creates an economy for digital archiving—a scarcity that publishers and consumers are willing to pay to overcome and that would sustain the common pool resources? Put another way, what kinds of exclusive benefits can be defined to induce parties to act in the public good and invest in digital archiving?

Over and over again in conversations with publishers, scholars, librarians, and academic administrators, we have found that one special privilege that would likely induce investment in digital archiving would be for the archive to bundle specific and limited forms of access services with its larger and primary responsibility for preservation. Although there is disagreement over the types of access services that would be desirable and permissable, the key phrase here is "specific and limited." User access in some form is needed in any case for an archive to certify that its content is viable. But "broad availability," to use the phrase that I quoted earlier from the proposed mission statement of prospective digital archive, goes too far.

Indeed, extended and complicated forms of access not only add to the costs of archiving, they also make publishers very nervous that the archives will in effect compete for their core business. We desperately need models of archival access that serve the public good; we do not need models that, in effect, set up archives as competitors to publishers, because publishers will find it very difficult to support them.

Secondary, non-competing uses might include aggregating for not-for-profit educational use a broad range of journals in the archive—a number of publications larger than any single publisher could amass—for data mining and reflecting the search results to individual publishers' sites. Another kind of limited, secondary use might be based on direct user access to the content, again for not-for-profit educational use, with "moving walls" of the kind pioneered in JSTOR. Still other possibilities exist for even further development. Files aggregated across publishers in the archives could serve secondary abstract and indexing publishers as a single source, both saving them from going to each and every publisher for the texts to index and enabling them to use computational linguistic and other modern techniques to improve their products. Source files might also be "born archival" at the publisher and deposited in the archive, from which they might then serve as the masters for the derivative published files that the publisher creates for its different markets. These latter two possibilities are not likely to emerge immediately, mainly because they would require intense negotiation among the interested parties, but they are suggestive of how a thoughtful, entrepreneurial, community-based approach to archiving might add incremental improvements that would actually lead to more dramatic transformations of the system of scholarly communications.

Much work still needs to be done to sort out what the right access models might be so that they attract the necessary ongoing flow of revenue to sustain the archives. But just as "broad availability" may be going too far on one side, so-called "dark" archives, in which a publisher can claim the benefit of preservation but yields no rights of access, goes too far on the other side. Finding the right balance is essential to moving forward in this complicated arena.

Conclusion

In a recent work entitled The Ethics of Meaning, Avishai Margalit observes that "shared memory in a modern society travels from person to person through institutions, such as archives, and through communal mnemonic devices, such as monuments and street signs." He might have added schools and universities to his list of institutions, and footnotes to his list of mnemonic devices. The task of sustaining these institutions and devices for memory is not an easy one, and is a burden that falls on us all collectively. None of us alone is responsible for that support, but we do, in concert with others, have a responsibility to make sure that

incentives are in place so that least some preserve the commons on which future scholarship and education so clearly depend. In other words, a division of labor is needed that is analogous to the complex division of labor that secures health care for the sick. "We are," Margalit writes, "collectively responsible to see to it that someone looks after the ill. But we are not obligated as individuals to do it ourselves, as long as there are enough people who will do it" (Margalit 2002; see also Appiah 2003).

References

Online resources checked March 30, 2004

Appiah, Kwame Anthony. 2003. "You must remember this." *The New York Review of Books* (50:4), March 13: pp. 35-37

Ayre, Catherine and Adrienne Muir. 2004. "The Right to Preserve: The Rights Issues of Digital Preservation." *D-Lib Magazine* 10 (March). Available at http://www.dlib.org/dlib/march04/ayre/03ayre.html.

Barton, Mary R., and Julie Harford Walker. 2003. "Building a Business Plan for DSpace, MIT Libraries' Digital Institutional Repository." *Journal of Digital Information* 4(2). Available online at http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Barton/

Beagrie, Neil. 2003. National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity. Washington, D.C.: Council on Library and Information Resources and the Library of Congress. Available online at http://www.clir.org/pubs/reports/pub116/pub116.pdf.

Cantara, Linda, ed. 2003. *Archiving Electronic Journals: Research Funded by the Andrew W. Mellon Foundation*. Washington, D.C.: The Digital Library Federation, Council on Library and Information Resources. Available at http://www.diglib.org/preserve/ejp.htm.

Crow, Raym. 2002. *The Case for Institutional Repositories: A SPARC Position Paper*. Release 1.0. Washington, D.C.: The Scholarly Publishing and Academic Resources Coalition. Available online at http://www.arl.org/sparc/IR/IR_Final_Release_102.pdf.

Dellavalle, Robert P., *et al.* 2003. "Going, Going, Gone: Lost Internet References." *Science* 302 (October 31): 787-788.

Dietz, Thomas, *et al.* 2003. "The Struggle to Govern the Commons." *Science* 302 (December 12): 1907-1912.

Grafton, Anthony. 1997. *The Footnote: A Curious History*. Cambridge: Harvard University Press

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science*, 162 (December 13): 1243-1248.

Hobbes, Thomas. 1651. *Leviathan*. Reprinted in 1934. London: J.M. Dent & Sons, Ltd.

Lavoie, Brian F. 2003. "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making." White paper published electronically by OCLC Research. Available online at: http://www.oclc.org/research/projects/digipres/incentives-dp.pdf.

Lawrence, Steve, *et al.* 2001. "Persistence of Web References in Scientific Research." *IEEE Computer* 34(2): 26–31.

Lessig, Lawrence. 2001. The Future of Ideas: The Fate of the Commons in a Connected World. New York: Random House

Library of Congress. 2002. Preserving Our Digital Heritage: Plan for the National Digital Information and Infrastructure Preservation Program. Washington, D.C.: Library of Congress.

Lynch, Clifford 2003. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL Bimonthly Report 226* (February). Available online at http://www.arl.org/newsltr/226/ir.html.

Margalit, Avishai. 2002. *The Ethics of Memory*. Cambridge: Harvard University Press.

O'Donnell, James. 2003. "Re: Vanishing Act." Electronic mail to liblicense-l@yale.edu, January 29. Available online at http://www.library.yale.edu/~llicense/ListArchives/0301/msg00118.html.

Ostrom, Elinor 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge: Cambridge University Press.

Ostrom, Elinor, *et al.* 1999. Revisiting the commons: local lessons, global challenges, *Science* 284 (April 9): 278-282.

Ostrom, Elinor, *et al.* 2002. "The Drama of the Commons." In *The Drama of the Commons*, Elinor Ostrom, *et al.*, eds. Washington, D.C.: National Academy Press, pp.

Schonfeld, Roger C., *et al.* 2004. "Library Periodicals Expenses: Comparison of Non-Subscription Costs of Print and Electronic Formats on a Life-Cycle Basis." *D-Lib Magazine* 10 (January). Available at http://www.dlib.org/dlib/january04/schonfeld/01schonfeld.html.

Tenner, Edward 2002. "Taking Bytes from Oblivion." *U.S. News & World Report* 132 (April 1): 66-67.

Waters, D., and J. Garrett, eds. 1996. *Preserving Digital Information: Report of the Task Force on Archiving of Digital Information*. Washington, D.C. and Mountain View, Calif.: The Commission on Preservation and Access and the Research Libraries Group. Also available at: http://www.rlg.org/ArchTF/.

Weiss, Rick. 2003. "On the Web, Research Work Proves Ephemeral: Electronic Archivists Are Playing Catch-Up in Trying to Keep Documents From Landing in History's Dustbin." *Washington Post*. November 24: A08.