

Arc



By Kevin M. Guthrie

Will We

in the
Digital
Age

There's a Will, But Is There a Way?

With the increasing use of electronic and network technologies for scholarly communication, there is considerable and justifiable concern in the academic community about the challenges associated with protecting electronic information for future generations of scholars and students. Before the development of these technologies, the long-term availability of scholarly research was ensured through a system in which libraries—especially academic and research libraries—purchased, stored, and preserved books and journals in paper. There is not yet an equivalent system in place to protect the electronic literature being published today. How can we be sure that such a system will evolve? Where will the resources come from to support it on an ongoing basis? Who will accept responsibility and accountability for such a system? These are just some of the challenges that lie ahead if the academic community's commitment to archiving is to make the transition to the digital age.

Kevin M. Guthrie is President of JSTOR.

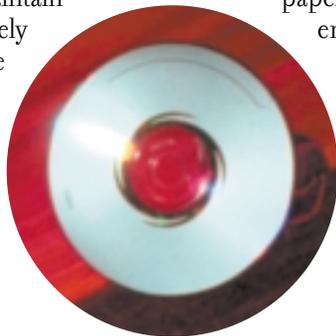
A Definition of the Term Archive

The word *archive* has different meanings depending on one's professional perspective. I am not using the term here in the traditional way that archivists refer to the collecting of manuscripts and letters. For the purpose of this discussion, to *archive* shall mean "to preserve and provide access to a collection of literature over time, without regard for how frequently these materials are being read or used."

Although the introduction of electronic technologies does not change this basic definition, it does alter the manner in which the tasks associated with preservation and access are carried out and it dramatically shortens the time period over which one can rely on the decisions made. In addition, this definition includes an economic component. Some portion of the collected materials will not be "economically productive": the total cost to digitize and maintain the information will exceed the amount of revenue that can be generated by making the materials accessible. Put another way, if archiving the information provided publishers or content owners with a reasonable return on the initial and ongoing investments required to maintain it, we could rely exclusively on the market to ensure that published literature remained accessible in perpetuity. But given past history and the nature of the use made of scholarly literature, this is not a safe assumption.

The Challenges of Preservation

Although preservation and access are interwoven in the electronic environment and in the above definition of the word *archive*, it is helpful to isolate them—to consider the two components separately—as we think about the impact of new technologies. The most dramatic opportunities presently associated with electronic archiving relate to access.



Archiving is not, and never has been, an issue fundamentally about technology; rather, it is about organizations and resources.

New technologies make it possible to store an item in one place and deliver it instantly across the globe. But electronic archiving also poses substantial challenges, most notably in the area of preservation. If electronic information is not actively and repeatedly updated, the technology and/or the formats used to store, read, and interpret it is likely to become obsolete. How can we ensure that materials in digital formats will be preserved and will remain readable when the technological environment is changing so rapidly? A discussion of this problem is far beyond the scope of this article. I can say here only that I do not believe there is a purely technical solution to it. Software cannot be written now that can predict the course of software development in the future.

In the present technological environment, then, the problem of preservation is ongoing and cannot, at any single point in time, be "solved" forever. But this is not a new development. Paper documents also decay, and in that sense, the challenge of long-term preservation has not been "solved" for the paper environment either. (It is just less obvious, especially to nonlibrarians, because the lifetime of paper documents, when properly sheltered, is so much longer than that of their electronic counterparts.)

One does not have to spend much time in a large library to find paper volumes and documents that cannot be used for much longer. It is often argued that even traditional preservation departments are underfunded. One 1999 study found: "Without question, the need for additional funds for preservation is urgent. The cost of preserving the nation's intellectual resources is high. Preservation is capital- and labor-intensive. Yet a look at the present funding support for preservation in research libraries reveals a discouraging picture. Preservation expenditures

in ARL libraries have been level since 1993."¹

So archiving is not, and never has been, an issue fundamentally about technology; rather, it is about organizations and resources. The degree to which students and scholars can now find and use the older literature as the building blocks of knowledge is a direct result of commitments made by libraries to store and protect the written word. Libraries have made and fulfilled these commitments because doing so has been a central aspect of their mission. Librarians have developed planning processes and organizational structures and have secured the resources to build and maintain storehouses of documents. As we look to the future and to electronic documents, who is going to carry forward the role that libraries have played, and what mechanism will they use?

The Existing Archiving System

In thinking about a future system for archiving electronic content, it is important to consider the motivations that led to the current system. To put it most succinctly, the development of the existing archiving system was not predicated primarily on a desire to archive; rather, archiving has been essentially a by-product of the need to provide access. If scholars and students at a college or university wanted to read a document, they needed to hold it in their hands, and the institution (typically the library) had to acquire the item, store it, and ensure its continued access. The mission and culture of the library profession, committed to preserving information and knowledge, thus operated to safeguard the information.

An important characteristic of this "archiving system" is that it emerged through countless local decisions. It was not a result of a central system-wide assessment, set of decisions, or investment focused on creating a trusted repository. Further, investing in the library had benefits that extended beyond holding the information content. The library was and is an important *place*, and the construction of large physical structures to house the accumulated knowledge lends credibility and prestige to the institutions that maintain them. Funding streams can be tapped, including fund-raising that

can associate donors with parts of the physical structure, helping to support the creation of these edifices of information. Once the huge fixed investments are made, the very small marginal cost of adding a volume to the physical collection supports the continuation of this pattern—until, of course, a library runs out of space, at which time the effort to raise the capital funds begins again.

The Electronic Journal and the JSTOR Example

What happens when access is unbundled from geography and ownership? It is no longer necessary to possess the electronic item physically—to actually hold the bits—in order to access and read the information. The items can reside on one server and reach any authorized user, anywhere in the world, who has a connection to the Internet. In such an environment, it simply does not make sense to duplicate the existing model, with its emphasis on local ownership. Individual colleges and universities are not demanding that content providers ship them electronic materials so that they can build a huge server and disk-storage infrastructure on campus to store, maintain, and deliver the items. Even if the publishers or content owners were comfortable with that approach, there are no longer local motivations for doing so.

So as documents, especially journals, increasingly become electronic, we are leaving the old model for archiving behind, without any replacement approach in operation.² One effort to address this problem is JSTOR, a not-for-profit organiza-

nization dedicated to building a central and trusted repository of back issues of journal literature (<http://www.jstor.org>). Yet though JSTOR concentrates on back issues, its electronic archive has a prospective as well as a retrospective component. For nearly all of the JSTOR titles, the prospective archival commitment is reflected by the “moving wall” that is included in JSTOR’s license agreements with publishers. As each year passes, an additional year of the journal becomes available in the JSTOR archive. Eventually, these additional “volumes” will be in the form of electronic journals that have been received directly from the publishers rather than the printed volumes that JSTOR currently scans and converts to electronic formats. Because most of the JSTOR moving walls are three to five years in length, these walls have not typically “run into” the electronic data now available on publishers’ sites. Consequently, JSTOR has come to be regarded by many as a retrospective archive of previously published paper journals rather than as a complete electronic archive with both a prospective and a retrospective responsibility. We at JSTOR expect that this assumption will change as the organization archives more content originally published in electronic form.

But even before that transition is made, there is a great deal to learn about the challenges associated with developing archives of electronic content. JSTOR has emphasized its archival mission from the outset and has attempted to demonstrate the benefits of approaching the archiving challenge by taking into ac-

count the needs of all key constituents in the scholarly communication process—scholars, students, publishers, and libraries.³ As time has passed, and the JSTOR archive has grown and become an actual resource as opposed to a vision, the argument that a central repository of electronic material might yield long-term savings even as it improves access is now supported by experience.

To offer an example, the JSTOR Arts & Sciences (A&S) I Collection presently includes 117 journals: about 7,700 volumes that would occupy about 1,200 linear feet of shelf space if stored in paper-bound format. Using calculations performed by Malcolm Getz and Michael Cooper,⁴ we can estimate the cost of storing and maintaining these volumes in open stacks at about \$16 per volume.⁵ Remote storage would cost approximately \$4 per volume. (These figures include only the capitalized cost associated with the construction and maintenance of the physical facilities.) Therefore, the total one-time cost for storing the A&S I Collection in open stacks would be approximately \$125,000. Adding the annual update to the moving wall would cost \$1,900. Remote storage for the complete run would cost approximately \$31,000 one-time and \$500 as each year is added.

The figures above do not include the costs of access, which are obviously significantly higher for remote access than for open-stack storage. What about these access and circulation costs? Before building the archive, JSTOR collected data on the use of ten important history and economics journals (which were

Comparative Costs for Holding and Providing Access to 7,700 Journal Volumes (Open Stack or Remote Storage versus JSTOR Participation)

	PAPER CONTENT		ELECTRONIC CONTENT
	Traditional Open-Stack Library	Remote Library Storage	JSTOR Medium Institution ¹
One-Time Cost ²	\$125,000	\$31,000	\$25,000
Annual Cost of Access ³	\$ 6,500	\$22,000	\$ 4,000

¹JSTOR relies primarily on the 1994 Carnegie Classification of Institutions of Higher Education to classify U.S. academic institutions. For a description of the JSTOR methodology, see <http://www.jstor.org/about/class.html>.

²For paper content, this includes building, shelf space, and capitalized maintenance costs. For electronic content, this is the one-time Archive Capital Fee (ACF) that a medium-sized library pays for participation in JSTOR. The ACF varies by institution type, with research libraries paying higher fees than small colleges (the fees range from \$45,000 for “Very Large” institutions to \$10,000 for “Very Small” institutions). For an explanation of JSTOR fees, see <http://www.jstor.org/about/ASI.pricing.html>.

³For paper content, this is the estimated annual cost of circulation. For electronic content, this is the Annual Access Fee (AAF) that a medium-sized library pays for participation. (AAFs range from \$8,500 for “Very Large” institutions to \$2,000 for “Very Small” institutions.) It should be noted that JSTOR participation has resulted in far more usage of the materials in electronic form than was predicted by the usage in the pilot program. Costs of circulation of the paper content for the level of JSTOR usage experienced at the five test-site libraries would be many times higher than the \$6,500 and \$22,000 figures shown here.

subsequently included in A&S I). In the paper format, the older journal volumes were consulted very infrequently. At five test-site libraries, a total of 692 uses were made of these journals over a three-month period.⁶ Extending that frequency of use over the entire 117-journal archive during the course of one year produces a conservative estimate of 6,500 uses of all of the journal volumes in A&S I for each of the five pilot libraries.⁷ Using data from the New York Public Library and Michael Cooper, we can estimate the cost of circulation at \$6,500 per year for open stacks and \$22,000 per year for remote storage.⁸

There is absolutely no doubt that the costs for an individual library to participate in JSTOR are substantially less—for smaller teaching institutions, more than ten times less—than the costs associated with storing and providing access to the paper journals. However, based on the feedback received from these libraries, the primary reason for JSTOR participation seems to be not the potential savings associated with central archiving but the benefits associated with providing better and more convenient access to the literature for faculty and students.

At these far lower costs, scholars' and students' ability to access the materials is greatly improved. In fact, usage of the older material has increased dramatically since it has been made available electronically in JSTOR. For example, it is projected that scholars and students at JSTOR-participating institutions around the world will print more than six million articles from the archive in 2001. It is also expected that approximately ten million searches will be performed.

This apparent tendency to justify JSTOR on the basis of the benefits of enhanced access as opposed to the potential savings associated with archiving is not by itself a bad thing. Nor is it surprising. For one thing, JSTOR is an extremely young or-

ganization, and libraries are appropriately conservative. They are not going to move quickly to realize the shelf space benefits soon after they sign up. They will want JSTOR to prove itself first. Further, the focus on access is consistent with the point made earlier: the primary motivation for an acquisition is its current usefulness, not its archival or future value to scholarship. Finally, and perhaps most important, librarians focus on the access benefits of JSTOR because, from a practical standpoint, this is all that the existing organizational and financial structures allow them to do. The benefits of newly available shelf space do not appear in any budget under their control. That budget resides somewhere else, such as under the control of the provost or president. Even for the provost, if the space available to the library is sufficient at present and for the near future, the economic savings is not a foremost consideration. It will be, though, when the time comes to build the library addition.

In many if not most cases, the only budget that the librarian can use to pay for an archive like JSTOR is the acquisitions budget. Assuming that the account used to purchase an item is related to the purpose of the purchase, reliance on the acquisitions budget to pay for JSTOR makes it a database, not an archive. In other words, by paying for JSTOR from the acquisitions budget, rather than from a combination of the acquisitions budget and funds that would normally be used to build new infrastructure for storage and archiving, institutions will have more difficulty recognizing, in economic terms, the cost savings associated with JSTOR participation. It is hard to take *all* costs into account.

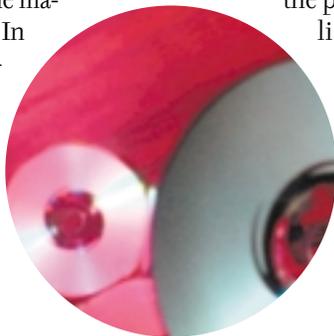
If JSTOR wants to make the archiving argument, to whom should it turn? Archiving is not the responsibility of the acquisitions librarian, who therefore cannot be expected to

act on that argument, however sympathetic he or she might be. Who is in charge of archiving at academic institutions? What budget in higher education institutions is dedicated to preserving this literature in the future? If electronic access is no longer tied to geography, and there are no local motivations for supporting an archive, how can we ensure that electronic archives will be established and maintained?

These questions highlight an early and important lesson of JSTOR. The argument that JSTOR offers an opportunity to realize true economic value—to reduce the cost of maintaining this material while offering enhanced access—is irrefutable. We can calculate the costs of the historical approach that has been taken and compare this approach against experience. But it is important to recognize that the accounting systems and organizational processes in academic institutions are not presently set up to realize that benefit. They do not make it easy to take all capital and operating costs into account when addressing such questions. Fortunately for JSTOR, the attractiveness of the back issues of journals included in the JSTOR archive—the “access value”—has been sufficiently high that library acquisitions budgets seem able to support the ongoing maintenance of the JSTOR archive.

Conclusion

If it is not possible to monetize these benefits in a situation where the costs and benefits can be estimated and projected with confidence, how are we going to address the huge electronic archiving problem? JSTOR's relatively few titles are a proverbial drop in the bucket. Are institutions going to be willing to contribute to centrally managed archives of important materials when the goal is specifically preservation, not access? The challenge that lies ahead, if little-used electronic materials are to be preserved for future scholars and students, is to address directly the motivations for investing in the resources to perform this function. How will institutions justify the investments in supporting or providing a centrally held electronic archive when there is no named building to point to, no beautiful library to show prospective



How will institutions justify the investments in supporting or providing a centrally held electronic archive when there is no named building to point to?

students, no addition to the volume count that leads to a ranking among the top libraries? Will colleges and universities be able to justify investments in archiving for its own sake?

If there is to be a transition from paper to electronic publishing, the archiving challenge will have to be addressed. This is not a problem that is going to be solved through a complex array of disconnected local decisions. Because electronic publication by its very nature depends on centrally held resources distributed widely through communications networks, in-

evitably some parties will be “providers” of an archive while others will be “beneficiaries” or “users.” We must determine ways to motivate the beneficiaries of the archive to support the costs of the providers. JSTOR will continue to test the question of whether support can be generated for a not-for-profit organization with archiving as its fundamental mission. But there are and will be other examples. One is the Andrew W. Mellon Foundation’s Electronic Journal Archiving Initiative, which has provided support to seven universities to develop

plans for archiving electronic content. These universities are working with a variety of publishers of scholarly content to test the technological, legal, and economic questions associated with archiving electronic content in a practical way.

This project and others like it will contribute to our understanding of what will be needed to preserve electronic documents. I hope they will also shed light on what motivates colleges and universities to protect and preserve scholarly literature for future users. The local motivations that have been the foundation of the current paper archive do not naturally generate the scale of resources that will be required to establish the more centralized model necessary for the preservation of electronic documents. There are examples, especially in the larger collecting institutions, of decisions made solely for the purpose of the long-term maintenance of the scholarly record. In those institutions, regular funding streams and even departments (e.g., preservation and conservation) have been established. These examples are the exceptions rather than the rule, but they demonstrate the kind of budgetary recognition and allocation that will be necessary if the problems of electronic archiving are to be overcome.

If the documents of today are to be preserved for tomorrow, the budgets of academic and cultural institutions may need a new line item in the digital age, a line item that transcends existing capital and operating budgets and addresses all the costs of maintaining the electronic record; perhaps it should be called “e-archiving.” Fortunately, one major benefit of the advances in information technology is that, from both a system-wide and an individual organization’s perspective, if costs can be spread over a significant number of institutions, an e-archiving budgetary commitment *can* be both less expensive and more effective than the regular investments already being made in the maintenance of the printed record. *e*

Notes

1. Jutta Reed-Scott, “Preserving Research Collections: A Collaboration between Librarians and Scholars” (1999), published by the Association of Research Libraries, the Modern Language Association, and the American Historical Association on behalf of the Task Force on the Preservation of the Artifact, <<http://www.arl.org/preserv/prc.html>> (accessed August 21, 2001).

2. For a general discussion of this terrain and a longer consideration of its specific application to JSTOR, see Kevin M. Guthrie, "Challenges and Opportunities Presented by Archiving in the Electronic Era," *Portal* 1, no. 2 (April 2001): 121, online version available at <<http://www.jstor.org/about/archiving.html>> (accessed August 21, 2001).
3. William G. Bowen, "JSTOR and the Economics of Scholarly Communication," October 4, 1994, <<http://www.mellon.org/jsesc.html>> (accessed August 21, 2001).
4. See Malcolm Getz, "Storing Information in Academic Libraries," mimeograph copy, October 17, 1994, and Michael Cooper, "A Cost Comparison of Alternative Book Storage Strategies," *Library Quarterly* 59, no. 3 (1989): 239-60.
5. This estimate assumes a twenty-five-year life and 10 percent discount rate.
6. Kevin M. Guthrie, "Revitalizing Older Published Literature: Preliminary Lessons from the Use of JSTOR," paper delivered at the PEAK Conference, March 23, 2000, online version available at <<http://www.jstor.org/about/preliminarylessons.html>> (accessed August 21, 2001).
7. Making these materials available in electronic format has dramatically altered the level of use upward. To estimate the cost of circulation, we used the far more conservative usage figures for print materials rather than the usage figures for electronic formats.
8. The New York Public Library estimated that closed-stack circulation cost \$1.92 per use. From this and other sources, we estimate circulation at traditional open-stack libraries to cost approximately half as much. See Bowen, "JSTOR and the Economics of Scholarly Communication."