

JSTOR: BUILDING AN INTERNET ACCESSIBLE DIGITAL ARCHIVE OF RETROSPECTIVE JOURNALS.*

By Richard De Gennaro

"No one in his right mind would do what you are proposing." That was the advice a prominent publisher gave William G. Bowen, President of the Andrew W. Mellon Foundation and founder of JSTOR, when he described the JSTOR concept to him in 1994.

History records many instances where significant breakthroughs were made by the naïve visionary on the periphery who does not know that what he is proposing "cannot be done." Bill Bowen had no idea how difficult it would be to bring his idea to fruition, but he was convinced that JSTOR would be a boon to scholars, publishers, and librarians. With the support of the Mellon Foundation he turned an impossible dream into it an idea whose time had come.

Originally a project of the Andrew W. Mellon Foundation, JSTOR is now an independent not-for-profit organization with a mission to help the scholarly community take advantage of advances in information technologies. JSTOR's initial objective is to develop a trusted archive of core scholarly journal literature, with an emphasis on the retrospective conversion of the entire backfiles of key journals. It is anticipated that in the future the objectives will be expanded and other related projects will be initiated.

In pursuing its mission, JSTOR is taking a system-wide approach, taking into account the needs of those involved in the field of scholarly communication: libraries, publishers, and individual scholars and students. This system-wide approach has required a willingness to reach compromise in order to accommodate the sometimes conflicting perspectives of JSTOR's constituents. Publishers needed terms that would preserve their copyright and their paper subscriptions. Librarians and publishers alike needed to define the limits of interlibrary loan, the definition of "authorized users" and the permitted use of

* Paper delivered at the 63rd IFLA General Conference, Amsterdam, August 31- September 5, 1997

JSTOR materials by those authorized users. With the resolution of these and other critical issues, JSTOR is now offering the first comprehensive collection of copyrighted retrospective journals in digital form to libraries and library users.¹

JSTOR's principal goal during the coming three years is to create a comprehensive database of the complete back files of core journals in 10-15 fields in the humanities and social sciences. The page images are scanned and bit-mapped to preservation standards at 600 dpi. The text is also optically scanned and the product is manually upgraded to 99.95% accuracy. This ASCII text serves as an index to the contents of the journals. The tables-of-contents are keyed and provide a convenient means of browsing the journals. In order to protect the publishers paper subscriptions, there is a 3-5 year gap between the back file and the current year. Each year another volume will be added to the database and archive.

Agreements have been negotiated with the publishers of some 52 core journals – the first 22 of which are now accessible on the Internet to Charter Participants. The JSTOR collection is being actively marketed to libraries in the US and Canada. The pricing schedule is based on the Carnegie Classification of Institutions of Higher Education with some modifications. JSTOR is committed to making a minimum of 100 journal titles available by the year 2000 as Phase I of a continuing effort.

In recent years there has been a good deal of discussion and speculation about the feasibility and usefulness of digitizing large quantities of printed books and journals in library stacks. There are some who believe that it is technically and economically feasible and intellectually desirable to digitize "everything", and there are others who think that the value of retrospective materials is limited and would not justify the high cost of digitizing and maintaining them. JSTOR takes the middle way. .

The JSTOR project is based on the conviction that it is intellectually desirable and economically feasible to digitize, maintain, and distribute a carefully selected body of core journals and other retrospective resources provided that the cost can be shared among a very large number of libraries. And it is also assumed that the cost savings to participating libraries will more than offset the fees that they pay. The Andrew W. Mellon Foundation has made grants of over \$4 million to establish JSTOR with the expectation that JSTOR will become an on-going, self-

¹ Consult the JSTOR website www.jstor.org for additional information about JSTOR including a list of titles completed and in progress, a pricing schedule, the library licensing agreement, a demonstration database, etc.

supporting, not-for-profit enterprise serving the needs of scholars, librarians, and publishers.

The JSTOR core journal collection is important to libraries for two reasons:

1. it greatly enhances user access, and
2. it provides significant savings in library space and operational costs. We will discuss them in sequence.

Enhanced User Access

The jstor database is to a library's collection of journals what a library's catalog is to its collection of books and journals, but whereas the catalog merely points to the book or journal, JSTOR indexes every significant word or phrase and instantly delivers the text to the user's desktop and enables printing on demand. This is what makes JSTOR a transforming scholarly resource.

In this easily accessible electronic form these core journals will acquire an importance that they did not have as bound volumes in library stacks. The JSTOR database is becoming a new and vital library resource. With desktop access, users will mine the contents of these journals in ways and to an extent that was simply impossible and inconceivable with the bound volumes. In addition, the JSTOR archive is complete while many if not most of the paper sets in libraries are incomplete with missing and mutilated volumes and defaced pages.

Digitized information tends to be more useful and valuable than print information. Experience with card catalogs and electronic catalogs has shown that users will prefer the convenience of the electronic catalog and neglect the cards. The same will be true of those journals that have been "jstored;" they will be used in preference to those that are only available in bound volumes. The use and value of the back files of paper journals will diminish as the number of digitized journals increases. This is why JSTOR is being so careful in the selection of the titles to be included in the database.

Savings in the Cost of Library Space and Operations

The JSTOR database meets preservation standards and faithfully replicates the contents of the journals including all advertising pages and membership lists, etc. But JSTOR is much more than a mere replication of the contents of the paper or microfilm copies of the original journals. As has been noted, the Internet-accessible JSTOR database is a totally new and unique research resource that will supersede the paper and microfilm sets of the journals it contains. To be sure, it is

essential that some major research libraries retain their bound sets of these journals for archival purposes, but most libraries could discard them or send them to secondary storage and free space in the bound journal stacks. Of course it may take some time for librarians and library users to gain enough experience with and confidence in JSTOR before taking this now seemingly drastic step. This decision will be much easier for that large number of libraries whose sets of these journals are largely scattered and incomplete.

JSTOR's original purpose was to save valuable stack space in libraries by substituting digitized versions of core journals for the bound sets that occupy large amounts of space in thousands of libraries. Saving space is still a major JSTOR goal and a compelling argument can be made for it if the cost of library space is viewed from an institutional perspective. Presidents, provosts, and chief financial officers are painfully aware of the cost of library space.² Because the cost of building and maintaining library space is usually not part of the library's operating budget, librarians have little incentive to factor the cost of space into their decision making. The cost of JSTOR will come out of the library budget, but the cost of the space it saves will accrue to the institution. Space is important to librarians because there is never enough, but they have little appreciation of its true cost. Space is still seen as a free good in most libraries, but there is some evidence of a trend toward charging space costs to library budgets. This has already happened at Harvard and a few other libraries.

Brian Hawkins, Vice President for Academic Planning and Administration at Brown University, has been questioning the financial viability of the traditional research library in a series of papers. He makes this sobering assessment in his most recent paper: "While the problems associated with the acquisition of new information are alarming, focusing on this set of costs masks the magnitude of the real problem. If we proceed with the library model as we have known it, the costs associated with storing and archiving the information will bankrupt our institutions of higher education."³

In the same paper Hawkins estimates the cost of physically housing a single volume at \$20, assuming new building costs at \$170 per square foot. In addition, he estimates annual maintenance costs at approximately \$1 per volume at Brown. Malcolm Getz in an unpublished 1994 study estimates the capital cost of construction of open shelf space at \$17.85 per volume.

² Brian L. Hawkins, "The Unsustainability of the Traditional Library and the Threat to Higher Education," Paper presented at the Stanford Forum for Higher Education Futures, The Aspen Institute, Aspen Meadows, Colorado, October 18, 1996. 19p. (Available from the Author).

³ Hawkins. p. 9

We estimated that there will be 6,400 volumes in the 100 titles in Phase 1. Using the standards in Leighton and Weber⁴ they will occupy 61 single faced sections of shelving (at 15 volumes per shelf and 105 volumes per section). The 61 sections would occupy 568 square feet at 9.8 square feet per section. Assuming \$200 per square foot for new construction in 1996, it would cost \$113,600 to construct the space occupied by the JSTOR volumes plus \$1 per volume or \$6,400 year to maintain it.

The purpose of these crude estimates is merely to highlight the space issue and to give some idea of possible savings. JSTOR will be encouraging and sponsoring more detailed studies on the impact of JSTOR on space and other library costs.

This discussion is limited to space and building maintenance costs in a single library. There are other potential savings in library operational costs such as savings in binding, preservation, repairs, retrieval, and reshelving. When these potential cost savings for a single library are multiplied by the number of libraries in a consortium, a state, in the United States, or the world, the numbers are truly staggering. Add to this the "have not" libraries that could gain access to this rich resource for the first time through JSTOR. The Andrew W. Mellon Foundation has a strong commitment to supporting higher education internationally. Its global perspective on the value of JSTOR accounts for its keen interest in and generous funding for this pioneering initiative.

Richard De Gennaro
JSTOR
70 Turtle Bay Drive
06405 Branford, CT
USA
Fax: 203 483-8364e-mail:
richard.degennaro@yale.edu

⁴ Metcalf, Keyes D., Second Edition by Leighton, Philip D. and Weber, David C. *Planning Academic and Research Library Buildings*, ALA, Chicago, 1985.p. 559 Table B.11, and p. 561, Table B.17.