

JSTOR, A Great Leap Forward in Electronic Journal Access



Wouldn't it be great if someone took those old, brittle, yellowing journals that take up so much library space and put them on the World Wide Web? Wouldn't it be great if they set it up so you

could print out from your desktop computer high-quality copies of exact replicas of the pages of those journals? And wouldn't it be great if they also made available a text-search capability?

This is not some futuristic fantasy—this is JSTOR. JSTOR, which stands for Journal Storage, has developed a database that provides Internet access to some of the most important scholarly journals in a variety of academic fields, including the mathematical sciences. JSTOR is a not-for-profit organization whose aim is to parlay digital technologies into new ways of preserving and making accessible scholarly journal literature. Certainly JSTOR is no panacea for the resource and storage problems facing academic libraries: At present JSTOR comprises only 76 journals, while statistics from the Association of Research Libraries indicate that a typical university purchased about 16,000 serials in 1997. Nevertheless, JSTOR represents an ambitious first step toward bringing journal literature into the digital age.

JSTOR started in 1994 as a pilot project, funded by the Andrew W. Mellon Foundation, to scan 750,000 pages from ten journals and make the pages accessible through computer networks. By 1995 JSTOR was established as an independent not-for-profit organization and had assembled the staff and computer resources needed to carry out a production scheme that currently processes ap-

proximately 100,000 pages a month. With the establishment of a new production facility at Princeton University, this capacity is expected to triple by the end of the year.

The JSTOR production process begins with journal acquisition. Full runs of journals are usually obtained from publishers, who often find it beneficial to cooperate with JSTOR, because they are unlikely to obtain further revenue from older back issues of journals and because storage on JSTOR helps to make their publications more widely accessible. The journals are sent to JSTOR's Ann Arbor office, which inspects the sets of journals to be sure they are complete and in good condition. The Ann Arbor office also establishes indexing guidelines for each journal so that there is consistency in how the bibliographic information is presented on JSTOR.

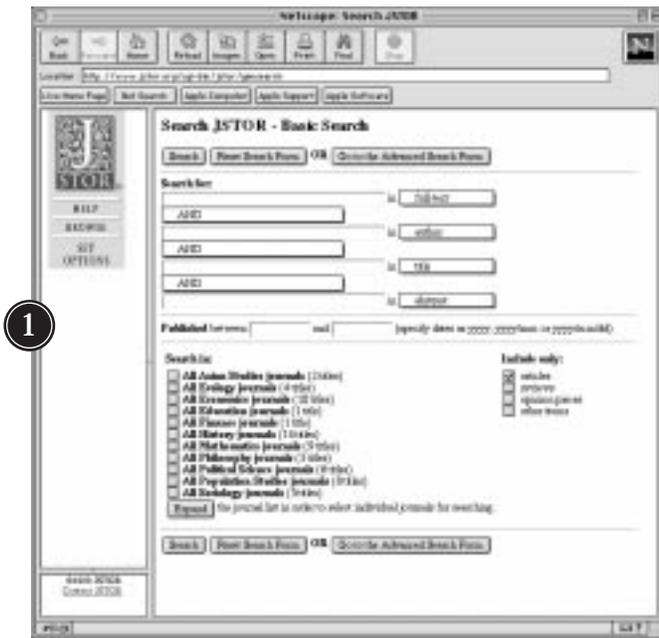
From Ann Arbor the journals go to an outside firm for scanning. The scanning is done at high resolution (600 dpi) and with meticulous care: The intention is that the quality of the scanned images is sufficiently high to meet the needs of whatever use the images might find in the future. Optical character recognition (OCR) software is used to convert each page into a text file, with an accuracy of 99.95%; these files permit text searches of the journals. Abstracts and bibliographic information are keyed in manually. At the end of this process, the data is stored on CD-ROMs and sent back to Ann Arbor, where quality control checks are performed before the journals are made available to subscribers on the Web.

JSTOR is now in phase I of its development, which aims to build a collection of at least one hundred important journal titles in a variety of fields by the year 2000. Decisions about which journals

to include in JSTOR depend on a number of factors. One is the amount of time a journal has been in existence: Libraries benefit more if JSTOR concentrates on journals with a large number of back issues rather than newer journals that take up less shelf space. Another is the impact of the journal, as measured by such things as the number of subscriptions and the number of citations its articles have received. JSTOR also consults with specialists in the various fields to decide which journals are the most important to include. The availability of full runs of back issues and the cooperation of publishers are considerations in many cases. In phase II, JSTOR will continue to expand its collection,

adding journals that are clustered in certain fields; it is likely that one of these phase II clusters will include mathematics journals. Libraries will have the option to choose clusters most suited to their needs.

While scanning the journals is a large part of the cost of producing JSTOR, there are other costs as well, such as indexing and quality control, developing software, maintaining and upgrading hardware, user support, and administration and oversight. JSTOR has developed an economic model based on the idea that the costs of developing and maintaining this archival resource can be spread among a large number of participating institutions. The fee these institutions pay to JSTOR can be justified by reduced storage costs, savings in long-term capital purchases, and improved service to library patrons. Currently, over 260 libraries in the U.S. and Canada participate. A new mirror site established at the University of Manchester will provide access to the JSTOR database in the United Kingdom. JSTOR has a sliding scale of prices based on the Carnegie Classification of the participating institution. There is a one-time "Database Development Fee", which guarantees participating institutions perpetual rights to information in the phase I archive. This fee ranges from \$10,000 for very small institutions to \$40,000 for large ones.



Counterclockwise from upper left, illustrations show (1) a search sequence starting with the basic JSTOR search screen, followed by (2) a portion of the Browse All Journals screen, (3) choice of available volumes for browsing of *Transactions of the American Mathematical Society*, (4) available issues of Volume 31 of *Transactions*, (5) contents of Volume 31, No. 3, and finally (6) a sample article from Volume 31, No. 3.



There is also an "Annual Access Fee" of \$2,000 to \$5,000.

JSTOR and the AMS have enjoyed fruitful cooperation from the start. Former AMS president Cathleen Morawetz serves on the JSTOR Board of Trustees, and AMS executive director John Ewing has had extensive discussions with JSTOR. Right now JSTOR is cooperating with *Mathematical Reviews* to develop a way to provide links from reviews in MathSciNet directly to the articles stored on JSTOR. This would add about 35,000 links. There is also a project under way to add the bibliographic data of about 3,000 pre-1940 mathematics articles from JSTOR to the MR database.

The JSTOR Web site is nicely designed and easy to navigate. There are various routes to find what you are looking for. You can browse through journal articles, clicking to flip from one page to the next or from one article to the next. You can do text searches on the full article text, author names, titles, and abstracts. You can also use an advanced search form to search with strings of text linked by Boolean operators. The images of article pages are stored as tiff files and load fairly quickly, and the quality of the images is quite good. A special application, called JPRINT, has been developed to facilitate printing articles from JSTOR. JPRINT can be downloaded from the JSTOR Web site; the instructions for doing so are ample and clear. JPRINT provides two options: fast printing at low resolution and high-quality printing that takes longer and has higher resolution. Samples of pages from *Proceedings of the AMS* printed on a laser printer were both perfectly legible, with the high-resolution copy being very close in quality to the printed journal. There is also a PDF option for viewing and printing.

JSTOR decided to offer only images of journal pages to be downloaded or printed, not the actual text. This decision was based on a number of factors, one being the large number of graphics, tables, and special symbols used in scholarly journals. There is currently no standard way to present these characters using typical Web browsers. In addition, it is expensive and difficult to use OCR to convert a scanned image into text with 100% accuracy (though the accuracy of OCR is just fine for generating a text file for searches). Making available 100%-accurate text files would have meant essentially republishing the journals from scratch.



6



4

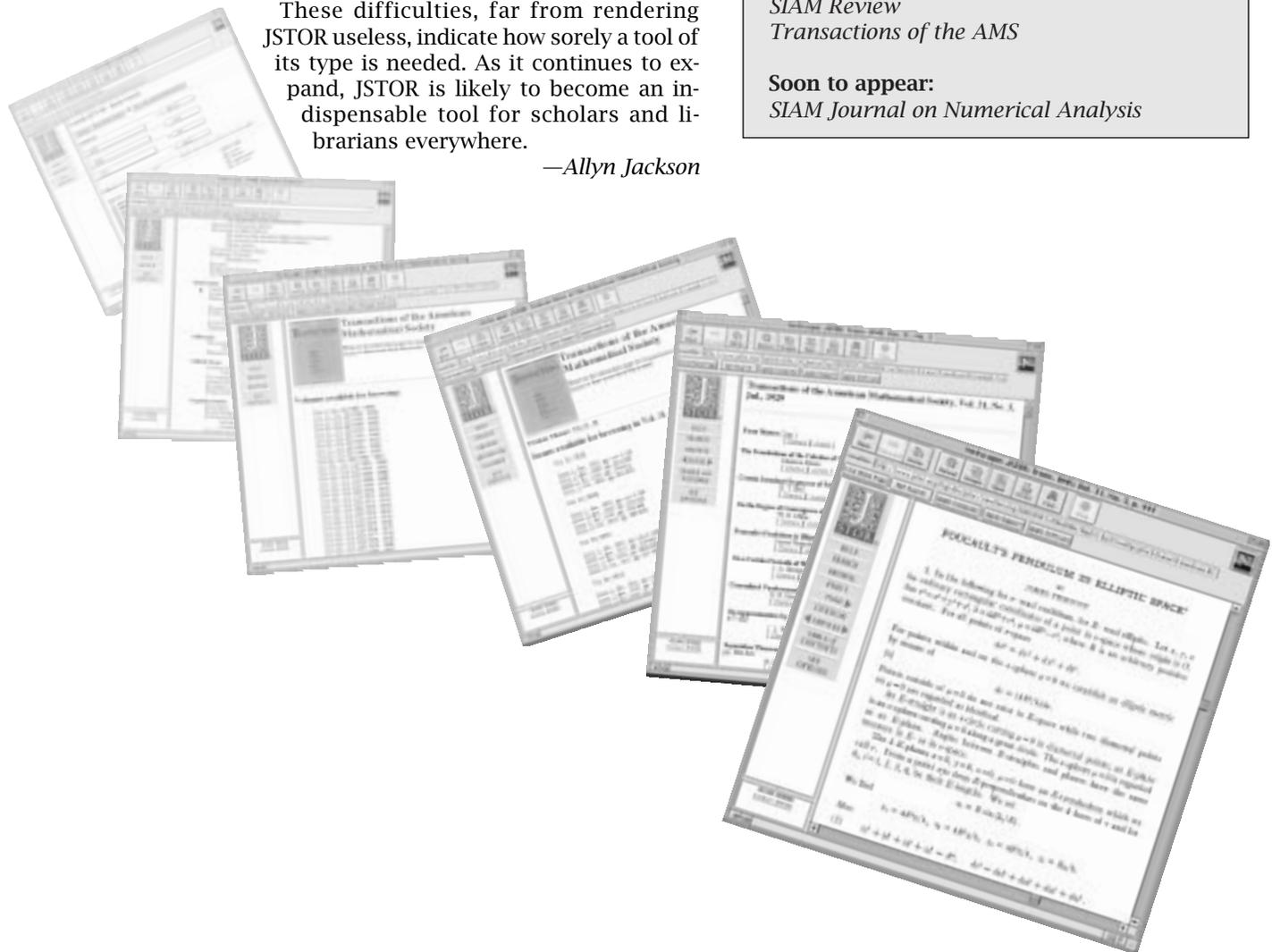


5

The primary drawback of JSTOR for mathematics (and probably for other fields as well) is that so few journals are included. Consider that about 575 journals are reviewed cover-to-cover by *Mathematical Reviews* (not to mention the hundreds more that contain individual items selected for review), and the 9 mathematical sciences journals contained in JSTOR look like small drops in a very large bucket. Another difficulty is that JSTOR does not include the most recent volumes of journals. It provides access to the complete backfiles of every journal available in the database, but has agreements with the publishers to remain from two to five years behind current issues. At the beginning of each calendar year JSTOR extends coverage of each journal by one year. This scheme, called a "moving wall", is designed to protect publishers from the threat of lost revenues. Thus, for example, the issues of *Transactions of the AMS* that are available on JSTOR start with the establishment of the journal in 1900, but only go up to 1992. One can link from JSTOR to the Web site of *Transactions*, where issues starting in 1996 are available online.

These difficulties, far from rendering JSTOR useless, indicate how sorely a tool of its type is needed. As it continues to expand, JSTOR is likely to become an indispensable tool for scholars and librarians everywhere.

—Allyn Jackson



JSTOR

188 Madison Avenue
New York, NY 10016
telephone 212-592-7345
fax 212-592-7355

World Wide Web: <http://www.jstor.org>

Kevin Guthrie, President
Margit A. E. Dementi, Associate Director,
Library Relations

Mathematics Journals on JSTOR:

Annals of Mathematics
Journal of the AMS
Mathematics of Computation
Mathematical Tables and Other Aids to Computation
Proceedings of the AMS
SIAM Journal on Applied Mathematics
Journal of the Society for Industrial and Applied Mathematics
SIAM Review
Transactions of the AMS

Soon to appear:

SIAM Journal on Numerical Analysis