

LETTER FROM THE PRESIDENT

Three times per year we have the opportunity, through JSTORNEWS, to report on recent JSTOR developments and to share with the scholarly community our experience in delivering a growing and important electronic resource. In this issue, you will find a significant amount of space dedicated to the topic of remote authentication. This is one of the most challenging problems facing any electronic resource provider. At the ALA meeting in July, Ira Fuchs, JSTOR's Chief Scientist, gave an overview of this problem and some of the potential solutions. This is such an important topic that we offer an abridged version of his talk here in the newsletter. I hope you will find it both interesting and educational. If you would like to learn more about remote authentication, please consult Ira's full paper, located at <http://www.jstor.org/about/remote.html> and also Cliff Lynch's "White Paper on Authentication and Access Management Issues in Cross-organizational Use of Networked Information Resources" located at <http://www.cni.org/projects/authentication/authentication-wp.html>.

Enthusiasm for JSTOR continues to build and we are energized by the challenge of meeting the growing expectations of the academic community. This enthusiasm is reflected in every possible statistic about our service. September and the beginning of a new academic year brought the highest monthly usage of JSTOR ever. There are now 344 participating libraries in the United States, Canada, and the United Kingdom, 97 Phase I titles signed on, and over 2.3 million pages available in the database. These are exciting times. Thank you very much for your support, encouragement and comments. Your feedback is important to help us continue to improve the services we provide.

Kevin M. Guthrie

JSTOR BEGINS WORK ON NEW GENERAL SCIENCES CLUSTER

JSTOR's new production office at Princeton University is rapidly filling up with back issues of *Science* and the *Proceedings of the National Academy of Sciences (PNAS)*, the first two titles to be included in the new General Science cluster. The development of the General Science cluster initiates JSTOR's Phase II plans. Unlike Phase I, which will include 15 disciplines and a minimum of 100 titles, we intend to offer future titles in smaller disciplinary clusters. These clusters will be offered separately from the JSTOR-Phase I database.

Science, the weekly peer reviewed journal published by the American Association for the Advancement of Science (AAAS), dates back to 1880. The cluster will also include *Scientific Monthly*, a sister journal published by the AAAS from 1915 through 1957. *PNAS* is published by the National Academy of Sciences and dates from 1915. Together, they comprise some 500,000 pages. Other titles for this cluster are still in negotiation.

"JSTOR represents a unique opportunity to create a comprehensive digital archive of *Science*," says Floyd E. Bloom, Editor-in-Chief of the journal. "Our objective in working with JSTOR is to ensure that future generations of scholars will have enhanced access to the complete back files of the journal."

According to Kevin M. Guthrie, President of JSTOR: "This collaborative effort between JSTOR, the AAAS, and the PNAS will be enormously valuable to the scholarly community. More than 100 years of important research in the sciences will be fully searchable and accessible via the World Wide Web. This will not only be a boon to science researchers, but also a treasure for studying the history of science and the evolution of scientific inquiry in the United States."

The General Sciences Cluster will be prepared, indexed and loaded at JSTOR's new Princeton-based production facility. This new production facility, which is administered through a special grant relationship with Princeton University, will be overseen by JSTOR's Princeton Production Coordinator Elizabeth Bennett. "Creating a searchable digital archive for *Science* and *PNAS* will be a new kind of challenge for JSTOR," says Bennett, "because the scientific literature contains more illustrations and color and is more complex in layout than much of JSTOR's Phase I content." As always, JSTOR's goal is to create a high-resolution, faithful replication of each journal page as it was originally printed, and to enhance that digital archive with indexing of authors, titles, abstracts, keywords and full-text to facilitate comprehensive and convenient access to the information in the database.

REMOTE AUTHENTICATION AND AUTHORIZATION FOR JSTOR

*adapted from a talk given by Ira Fuchs
at the JSTOR participants' meeting, ALA Meeting, June 28, 1998*

One of the most vexing issues facing providers of digital content, libraries and users today is how to provide convenient portable access to electronic resources. As members of university communities have discovered the importance of networked information, they have come to expect access at any time from any place. Currently, many content providers control access through the use of IP addresses, an approach that makes sense when users are connecting from on-campus. But reliance upon IP addresses does not work when users are at home, or when they are on leave, on sabbatical, or traveling. The key constituents in the scholarly communication process — content providers, campuses licensing the resources, and users — all need reliable authentication and authorization schemes that allow remote access and that can distinguish legitimate users from others seeking unauthorized entry. Each of these communities has specific interests that need to be balanced in any authentication solution.

Campuses require authentication schemes that are affordable, that perform reliably and are reasonably secure, that are relatively easy to install and maintain, and that scale effectively for all members of the community whether they are on- or off-campus. Content providers like JSTOR need scalable and maintainable systems to preserve the integrity of their usage licenses and to protect their resource from unauthorized access while simultaneously providing access for legitimate users. Users want convenient schemes that don't impede access, that offer a single sign-on rather than requiring separate login processes for each of many resources, and that enable access to electronic resources from anyplace at anytime. Presently there is no solution available that balances these needs perfectly. Briefly described below are three classes of alternative solutions and the primary advantages and disadvantages of each approach.

Proxy Servers: A proxy server is a program running on a computer located within the authorized campus domain (as defined by IP addresses) that acts as an intermediary be-

tween an end user and a target web site. That is, the normally point-to-point contact between browser and web site server is diverted into a two-hop channel: browser to proxy server, then proxy server to Web server, then back on the same path. Because user requests come to the content provider from a proxy that has an IP address within the authorized campus domain, the user's request can be accepted. There are two kinds of proxy servers: mechanical proxy servers and application proxy servers. Mechanical proxies are set up in such a way that they are transparent to the content provider's web server, but that requires that users make special modifications to their browsers in order to use them. This configuration process need be done only once, thereafter the proxy server is invisible to both end users and to content providers. Unfortunately many users do not want to go to this trouble. This approach also may not be acceptable because it is difficult to implement for library administrators responsible for many machines.

Application proxies are set up to be transparent to the end user, but they are not transparent to the content provider's web server. Like mechanical proxies, application proxies mediate between users and web servers, but they do so by dynamically rewriting URLs on a local proxy so that the end-user machine never interacts with the target web-server. The most serious disadvantage of this approach is that the proxy's protocol translation may cease functioning if the content provider employs tools such as Java, JavaScript, or complex HTML. Even if the application proxy does work, any change to the content provider's web page could cause the proxy to malfunction. Because of this fact, maintaining application proxies can be extremely complex.

That points to a general disadvantage of proxy servers; they can be expensive for campuses to maintain. As the number of users grows, the campus needs to face issues related to scaling the server and staffing its maintenance in such a way that it does not become a bottleneck in the information management system. The cost of the computational resources, maintenance and support costs would therefore be likely to climb in proportion to the number of information services. It would be preferable, then, for an institution to have a single authentication proxy for many resources, rather than separate proxy servers for each resource.

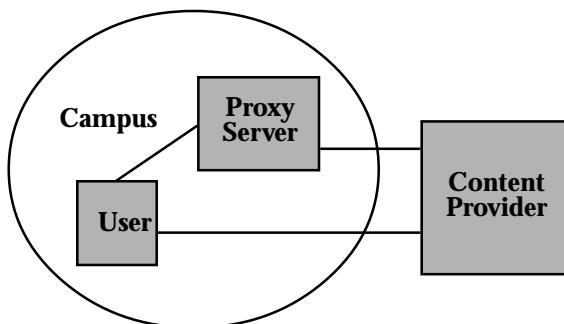
Credential Based Approaches: In the credential based approaches, the user interacts directly with the content provider rather than through a proxy operated by the university. The most well-known case, and one prevalent for many net-based resources, is one in which the user offers his/her username and password in order to logon to an information service. This access scheme has many obvious advantages. The user community is already familiar with the idea of logging onto a service. Users are already aware that they need to exercise care in selecting, securing, and occasionally changing their passwords. Unlike certificates (see below) which need to be installed on specific machines, users understand that they must remember their own passwords, but if they do, they can gain access to information services from any machine connected to the Internet. There are also disadvantages with this access scheme. The access system does not scale effectively because every con-

tent provider needs to enact a system for checking usernames and passwords. The approach complicates life for the user because she or he will have to keep track of distinct user names and passwords for each content provider they use.

Another, more complex credential-based access approach employs digital certificates. Certificates verify both the authenticity of the sender of the message as well as the integrity of the message itself. X.509 certificates provide all members of the user community with an international standard form for presenting their credentials. Once the certificates have been installed in a user's browser, the access system is invisible to the end user. Unlike using direct logons for access, user access to multiple sites is seamless. And security here does not require that the content provider verify or otherwise secure the user's name and password. There are again several disadvantages. Universities will need to manage the distribution of certificates. Users will need to install certificates in every machine that they use for access to licensed content. The scheme will also require the cooperation of content providers in accepting certificates.

Proxy/Credential Hybrid: In this final access scheme, the user gains access to the information resource through an application proxy that authenticates the user, authorizes the use of the information resource, forwards the approved credentials to the information provider, and then "gets out of the way." (JSTOR has implemented a script to perform this function. If you want to learn more about it, contact jstor-info@umich.edu.) The content provider approves the user's credentials by issuing a "ticket" that typically contains an expiration time. Users who want to sustain access after that time will need to acquire a new ticket. Using this Proxy/Credential Hybrid, campuses will be able to avoid the cost of managing certificates. In addition, apart from having to enter a username and password, the access scheme is invisible to the end user and has less overhead than the application proxy because the actual interactions are performed without support from the proxy itself. As a result, the resource requirements for the proxy will be more modest and there will be less degradation of performance.

Proxy/Credential Hybrid



The user gains access to the content provider through an application proxy that authenticates the user, authorizes the use of the information service, and then forwards the approved credentials to the content provider. Once the user's credentials have been approved, the user will interact directly with the content provider.

Conclusion: One conclusion from this review is clear. Campuses must establish authentication schemes. Without campus-administered authentication, even the most straightforward access schemes exclude legitimate members of the campus community and run the risk of jeopardizing site licenses as well as participation in inter-campus agreements that may offer significant financial benefits. It is simply not possible for resource providers to remain current on which students are registered at an institution or which faculty remain on staff. Campuses will favor schemes that are affordable, that are relatively easy to install and maintain, that perform reliably and are reasonably secure, and that scale effectively for all members of the community. Although there remain some compromises, any of these three approaches can be employed to offer remote access to authorized users of campus resources. We will continue to monitor the development of technology in this area in an effort to provide the best possible solutions for JSTOR users.

JSTOR REVIEWS AND ADJUSTS FEES AND CLASSES

At JSTOR we recognize that we operate in a dynamic environment. Every activity undertaken and every decision made is driven by our mission to increase access to and preserve scholarly material while at the same time saving library resources over the long run. Decisions are made with the knowledge that they will need to be revisited and revised as necessary. One aspect of our activities that merits regular reevaluation is pricing.

Our goal is to be certain that our fee structure is fair. We pursue a value-based approach that seeks to match the amount institutions contribute to the value they will receive from participation. During the summer of 1998 we conducted a thorough review of our fee structure and library classification policy. Statistics from more than 1500 universities and colleges, both JSTOR participants and non-participants were input and analyzed. We looked for similarities and differences among them, and ways in which they fall into groups. The statistics used were some of the traditional library collection standards such as the number of volumes in the collection and the size of annual serials budgets. These were compared with the JSTOR usage statistics that had been gathered over the first full academic year of library participation. What we found during this review was that some of our classifications needed to be adjusted, and that the Database Development Fees for some classes needed to be changed.

JSTOR's classification of libraries is based on the 1994 Carnegie Classification of Institutions of Higher Education. This classification "groups American colleges and universities according to their missions", and allows JSTOR to vary fees based on the degree to which institutions regard research as a fundamental aim. Using Carnegie classes has the added appeal that they are well understood in the academic community, and the assignment of individual institutions to the various Carnegie categories is done by a respected outside party. On the basis of Carnegie

classifications JSTOR initially created four groups of institutions: Large, Medium, Small and Very Small. When we re-examined the libraries in these four groups it was immediately apparent that Research I institutions stand out in a class very different from all others. The value they are getting from JSTOR participation is significantly greater than any other group of libraries, and their collecting patterns are also very different. It was felt that these differences should be reflected in the fees they pay. Hence, beginning January 1, 1999, there will be a new JSTOR class called Very Large which will include all Research I institutions. The Database Development Fee for this group will be \$45,000. To further distinguish the Research I libraries from all others, the DDF's for several other classes were decreased, effective August 1, 1998. The table on the right illustrates the changes made to JSTOR's fee structure and methodology for classifying libraries.

The DDF decreases are certainly good news for libraries that are still deciding about JSTOR participation. They have the opportunity to sign on to JSTOR for a lower fee. One librarian, discussing his institution's recent decision to participate in JSTOR, said that the DDF decrease was what made it possible for them to come in now.

The decreased fees also affect some JSTOR participating libraries. We appreciate the support JSTOR received from libraries that decided to participate during our early days; they also benefit from these changes. As soon as the new fee structure

was announced, JSTOR offered a rebate to those libraries that had come in at the previous rates. In accepting the rebate, most libraries requested that JSTOR apply the money to future Annual Access Fees. One librarian suggested that JSTOR keep the rebate and instead add more journals to the database. This continued enthusiasm and support for JSTOR is a welcome motivation for us all to do more.

Summary of JSTOR
Fee Changes and Classification Adjustments

Price Adjustment	Carnegie Class	Institutions in Class	Database Development Fee	
			Previous	Present
<i>New Class</i> -Very Large	Res I	88	\$40,000	\$45,000
<i>Price Adjustment</i> -Large	Res II, Doc I	86	\$40,000	\$35,000
<i>Price Adjustment</i> -Medium	Doc II, MA I	495	\$30,000	\$25,000
<i>Group Downgrade</i> -to Small	MA II	260	\$30,000	\$20,000
<i>Group Downgrade</i> -to Very Small	BA II	471	\$20,000	\$10,000

<http://www.jstor.org>

COMMENTS

I've been romping around in back issues of *Annual Review of Sociology* and *Annual Review of Ecology and Systematics*, browsing to my heart's content. It looks terrific. It's a wonderful service you're providing to the scholarly community and to the libraries.

Sam Gubins, President and Editor-in-Chief, Annual Reviews

JSTOR has done an outstanding job with mathematics. Retrieval of articles in these journals is now not only easier, but the search options greatly expand our ability to use older journals. Perhaps the greatest impact however has been that it has allowed us to keep our branch library within the math building. We ran out of space for journals and books, having exceeded the load limits per square foot. Because our faculty was satisfied with the JSTOR online system, we were able to move older issues of such space filling journals as the Transactions of the A.M.S. Now we can afford to continue to have books and journals located in our library for another 4 or more years. Please, please add more math journals!

Donald E. Marshall, Professor, University of Washington

I am really impressed. I just want to say that you all run a really terrific service. What prompts this outburst is that you are the only one from whom I can get monthly usage statistics on the first day of the new month. Sounds like a small item, but to me it is symbolic of a well thought out system, and that is important. Similarly you are the only one that gives us regular and timely updates on new info available. It's always a pleasure to hear from and communicate with you. Thanks for your good work.

William Tubbs, Librarian, Miller Library, Washington College (Chestertown, Maryland)

JSTOR FACTS

1998 Statistics
January - September

Total Accesses: 2,934,031
 Searches Performed: 727,707
 Articles Searched: 541,426
 Articles Printed: 235,825
 Pages viewed: 1,157,462

Total issues currently available: 12,802
 Total full-length articles available: 140,766
 Total articles (incl. letters, etc.): 388,026
 Total pages currently available: 2,313,483

Number of participating institutions: 344
 Number of participating journals: 97
 Number of participating publishers: 55