Scholar's Forum: A New Model For Scholarly Communication

By Anne M. Buck, Richard C. Flagan and Betsy Coles
California Institute of Technology, Pasadena, CA, March 23, 1999

> "A revolution must be wrought in the ways in which we make, store, and consult the record of accomplishment…It is not just a problem for the libraries, although that is important. Rather, the problem is how creative men think, and what can be done to help them think. It is a problem of how the great mass of material shall be handled so that the individual can draw from it what he needs-instantly, correctly, and with utter freedom. Compact storage of desired material and swift selective access to it are the two basic elements of the problem."
>
> Vannevar Bush,
> *Science Is Not Enough*, 1967

I. BACKGROUND

Scholarly journals have flourished for over 300 years because they successfully address a broad range of authors' needs: to communicate findings to colleagues, to establish precedence of their work, to gain validation through peer review, to establish their reputation, to know the final version of their work is secure, and to know their work will be accessible by future scholars. Eventually, the development of comprehensive paper and then electronic indexes allowed past work to be readily identified and cited. Just as postal service made it possible to share scholarly work regularly and among a broad readership, the Internet now provides a distribution channel with the power to reduce publication time and to expand traditional print formats by supporting multi-media options and threaded discourse.

Despite widespread acceptance of the web by the academic and research community, the incorporation of advanced network technology into a new paradigm for scholarly communication by the publishers of print journals has not materialized. Nor have journal publishers used the lower cost of distribution on the web to make online versions of journals available at lower prices than print versions. It is becoming increasingly clear to the scholarly community that we must envision and develop for ourselves a new, affordable model for disseminating and preserving results, that synthesizes digital technology and the ongoing needs of scholars.

In March 1997, with support from the Engineering Information Foundation, Caltech sponsored a Conference on Scholarly Communication to open a dialogue around key issues and to consider the feasibility of alternative undertakings. A general consensus emerged recognizing that the certification of scholarly articles through peer review could be "decoupled" from the rest of the publishing process, and that the peer review process is already supported by the universities whose faculty serve as editors, members of editorial boards, and referees.

In the meantime, pressure to enact regressive copyright legislation has added another important element. The ease with which electronic files may be copied and forwarded has encouraged publishers and other owners of copyrighted material to seek means for denying access to anything they own in digital form to all but active subscribers or licensees. Furthermore, should publishers retain the only version of a publication in a digital form, there is a significant risk that

this material may eventually be lost through culling little-used or unprofitable back-files, through not investing in conversion expense as technology evolves, through changes in ownership, or through catastrophic physical events. Such a scenario presents an intolerable threat to the future of scholarship.

II. A NEW MODEL

The scholarly community has sufficient expertise and incentive to collaborate on the design of a new model for scholarly communication that takes advantage of networking technology and extends the traditional benefits of print journals. Such a model, while facilitating the exchange of findings and the preservation of the scholarly record, must also:

1. Support peer review and authentication
2. Support new models of presentation incorporating network technology
3. Permit "threaded" online discourse
4. Adapt to varying criteria among disciplines
5. Assure the security of data
6. Reduce production time and expense
7. Include automated indexing
8. Provide multiple search options

Incorporating these features, many of which already exist in some form, into a practical model requires visionary leadership, investment by sponsoring institutions, and collaboration among developers and those groups or individuals who volunteer to partner in specific experiments.

III. TRILATERAL PARTNERSHIP

Three major entities come together to implement this model:

- CONSORTIUM OF UNIVERSITIES
  Universities collaborating as a group or possibly under the aegis of a supra-university body such as The American Association of Universities, form a Consortium that assumes responsibility for: maintaining the servers for the Consortium, developing and maintaining operating standards and protocols, and supporting the preservation of the Consortium's scholarly record.

- PROFESSIONAL SOCIETIES
  Within the various disciplines, professional societies, committees, and working groups continue to establish journals with editorial boards that are commissioned to review and validate work submitted by authors for final publication. Societies retain the power to publish and sell their journals in print or non-networked electronic formats such as CD-ROM or DVD-ROM; for the foreseeable future, many readers are likely to prefer receiving subscriptions as they do now.

- AUTHORS
  Supported by easy-to-use inputting protocols and standards, authors perform their own technical writing, copy editing, document formatting, etc., or else contract for these services from technical writing consultants (see Section V, Document Preparation Services). They may submit preliminary findings or preprints to the preprint database, or finished work directly to an editorial board for formal review.

IV. DOCUMENT DATABASE

The centerpiece of this proposal is a document database that incorporates and builds on important features derived from Paul Ginsparg's highly successful physics preprint server. Begun in 1991 and today comprising nearly 100,000 records in physics and related disciplines, xxx.lanl.gov demonstrates the viability of a large electronic archive that supports alerting services, automated hyperlink referencing, indexing, searching, and archiving. The proposed model also incorporates Ginsparg's recently developed plan to create an "intermediate buffer layer" overlaid on the raw preprint database and containing papers that have been subjected to a formal peer review. Such refereed papers may be aggregated into one or more journals that may exist at the buffer level. This heterodoxical approach opens the possibility for authors to establish their reputations simultaneously in a variety of related fields. Further value is added by shortening the reader's path to the certified version of a paper and by using links to point the reader back to the database of preprints.

V. UNIQUE FEATURES

EDITORIAL BOARDS
Editorial boards obtain permission from the Consortium to create and support a journal on Consortium servers. Following the tradition of confidentiality, a board determines whether a paper merits inclusion; it recommends revisions to authors; it considers authors' responses and rebuttals to referees' critiques; and ultimately accepts or rejects the work. An editorial board may also establish standards for document preparation. Revised versions that are placed in the preprint server receive a "version stamp". Eventually a "watermark", indicating final acceptance, is applied to the certified version that will be retained in all permanent archives maintained by the Consortium.

Consortium editorial boards are not granted exclusivity, i.e., any paper may be accepted for inclusion in multiple "journals". In addition, the editorial boards may not exclude a paper based on "prior publication" in the preprint server or elsewhere.

CONFERENCE PROCEEDINGS
Professional groups or associations, whether or not they are officially recognized societies, such as the organizers of a technical conference, workshop, or symposium, may apply for recognition by the Consortium as an editorial board establishing a new title within which reviewed and accepted works will be archived. A useful service for such groups may be to

invite them to enter abstracts or preprints into the preprint server before the conference and then aggregate the reviewed and accepted conference papers in an electronic conference proceedings, thus obviating the expense of publishing and distributing them in paper.

DOCUMENT PREPARATION SERVICES
Authors may require considerable assistance in preparing manuscripts that meet editorial boards' submission standards. In this model, the Consortium supports a directory of independent technical writers and editors with expertise in a variety of fields. These consultants may apply for inclusion or be recommended by an editorial board. The Consortium may also devise a procedure for certifying those who offer to provide document preparation services on a contract basis to authors.

COPYRIGHT
Authors or universities retain copyright according to institution policies. A mechanism at the input level requires authors to grant a limited, non-exclusive license to the Consortium. This agreement grants the right to provide unlimited access to all work in either preprint or archival servers for non-commercial purposes for the term of the copyright. Authors may grant limited-use licenses for their work to other not-for-profits or commercial entities, for which they may receive compensation, as long as such agreements do not infringe upon any rights previously assigned to the Consortium.

ACCESS AND REGISTRATION
All readers upon entering the database are notified that the contents represent protected intellectual property; readers must acknowledge acceptance of the terms of use. Readers may print out papers as well as make personal electronic copies for personal use, and they may forward URL's. However, readers must register to contribute to discussions and grant the Consortium limited, non-exclusive license to their contributions. Copyright is retained by the author of comments.

THREADED DISCOURSE
The model supports threaded discourse based on the work of researchers from Rand and Caltech to create a HyperForum. Colleagues may participate in dialogue on findings, however, anonymous comments will not be accepted.

The preprint server with its threaded discourse permits editorial boards not only to follow comments from the field, but also to identify important work and invite submission for review leading to inclusion in a journal. Of particular value is the opportunity for an editorial board to incorporate into their journal work usually associated with another field but of special interest to theirs. Concomitantly, this feature overcomes the need to require authors to prepare a new version of existing work.

SERVER MAINTENANCE AND SYSTEMS SUPPORT
As an extension of their on-going responsibility for assuring access to and security of scholarly publication, university libraries within the Consortium will be made responsible for maintaining the servers containing preprints, journals, proceedings, and abstracts of accepted

papers. Libraries may also provide system support through their own Information Technology groups or in partnership with campus IT organizations.

PRESERVATION AND ARCHIVING
Key responsibility for preserving the contents of the Consortium servers (in electronic formats) and archiving the final refereed papers (on 200-year secure paper) remains the domain of university libraries. The Consortium will designate multiple repositories among its members of both the electronic and print archival record to lessen vulnerability to catastrophic loss. Paper copies will be made, bound, and stored until an economical and completely reliable means for converting electronic data becomes available.

VI. STANDARDS AND PROTOCOLS

The overall cost effectiveness of the model requires the installation of data input and management protocols at the systems level. This platform enables individual authors to edit their own work while assuring that submissions are uniform throughout the database. While many relevant standards and authoritative systems already exist, the Consortium must invest in original work to design and develop an easy-to-use platform to support a variety of activities within the model:

DATA/TEXT ENTRY:
Authors will be able to work in their accustomed environments when writing and editing. Consistent document formatting will be ensured by the development of LaTeX macro packages and Microsoft Word™ styles or other data entry tools for authors to use. A limited number of standard methods for marking up equations will be supported initially (e.g., MathML, Chemical Markup Language, LaTeX). Authors will be encouraged to submit their work (and resubmit it after editing) through a simple web interface that will also allow them to provide basic metadata.

SUBJECT AND NAME IDENTIFYING
Subject identifier codes will be drawn from established, authoritative systems for contextual relevance. The system will support multiple identifiers/thesauri for a given work, and works will remain linked to the relevant thesauri so that as thesauri are refined and updated, the subject terms assigned to a paper will be automatically updated as well.

Form of name for authors and co-authors will be controlled as in traditional library authority-control systems. This will eliminate one of the major frustrations of current online systems where works by the same author are not collocated because the name or the form of name differs. Names will be dynamically linked as well; if the form of a name is updated, the linked papers will reflect the change. Subject and name authority systems will support the provision of cross-references.

RESOURCE DISCOVERY
Subjects and names as well as other metadata and full text will be searchable using the best available technology, including keyword and phrase searching, Boolean operators, proximity,

truncation, and relevance ranking. It will also be possible to browse the archive by subject term, author name, or chronologically.

DATA SECURITY

The integrity of the data in the archive can be supported initially by simple techniques like hashing. As digital watermarking standards evolve, this capability will be added. Current work at the World Wide Web Consortium (W3C) on versioning control and digital signatures shows promise in allowing the ability to identify and retain different versions of papers and to authenticate material that is served. Automatic mirroring technology applied by members of the Consortium will ensure against catastrophic data loss.

OBJECT IDENTIFIER SYSTEMS

Electronic publications must be identified uniquely and persistently. They must also be permanently locatable through a mechanism far less brittle than URLs. The Digital Object Identifier (DOI) is a step in the right direction for identifiers, although it does not provide all desired capabilities. Further work will be required in this area. CNRI's "Handle System" implementing Uniform Resource Names (URNs) shows great promise for generating persistent and ubiquitous document locators; developing a means for incorporating this capability is crucial to this model.

ARCHIVING

Papers accepted for the journal servers will be converted to XML (Extensible Markup Language) as an archival format. They will continue to be available for retrieval as HTML, postscript, pdf, or whatever format of choice may appear in the future. Metadata will be encoded in the Resource Description Format (RDF), an XML-based format for metadata which recently became a W3C standard. The use of XML for both data and metadata will ensure that either or both can be converted to other formats if the need should arise, in essence "future-proofing" the archive.

DIALOGUE MANAGEMENT

The Caltech Hyperforum model serves as a practical basis for providing "threaded" dialogue among registered participants. This model may be upgraded to provide more robust archival of comments using XML as a data format. This change would enhance retrievability of comments by various criteria as well.

MULTIMEDIA FORMATS

Future journal articles may comprise more than the present text and equations with perhaps a few illustrations. The multiplicity of multimedia formats that may be included in compound documents provides a fertile area for work on archiving standards and processes. The system will support current important formats and will provide for conversion as media formats change.

ANNOUNCEMENT SERVICE

The system will provide a current-awareness service allowing users to register and maintain "profiles" of interests through a simple web form. The addition of new papers on either the preprint server or the journal servers will trigger notification to interested users.

OVERALL DATABASE MANAGEMENT & COMMUNICATION

The active electronic journal model provides a very interesting challenge in the area of maintaining robust links between the journals and the archive, and also between the archive and the authority control (subject and name) systems. Extensible linking technologies such as Xlink and Xpointer, currently in development by the W3C, provide advanced linking functions, including the ability to: maintain databases of links separately from the documents or metadata they reference, address locations within documents without modifying the documents themselves, and assign "roles" to links that specify their function in the system and aid users in resource discovery. Work is this area will greatly enhance the robustness and usability of the system.

VII.   NEXT STEPS

The success of this model depends critically on winning the support of "champions" from the research community and attracting participants in initial experiments who are likely to come from emerging areas of research that have not yet had their journals published either commercially or by professional societies. Partnering benefits such groups by allowing them to leverage Consortium resources to announce their findings economically and to a broad audience. Potential partners may be identified by technical reference librarians through their contacts with faculty, students, and researchers on their campuses, or by searching the recent journal literature.

Before this is accomplished, research universities must assemble a Consortium to support the development and implementation of this model. The Consortium must assign lead participants from university IT departments, libraries, and faculty; identify and define elements of cost and develop a budget; establish a production schedule; develop underlying systems, standards, and protocols to enable champions, editors to create new journals; and attract funding from within the Consortium and from external sources.

CONCLUSION

The Scholar's Forum presents a unique approach that integrates in one conceptual model the elements of scholarly communication beginning at the author's keyboard and ending in the library's archives. By combining uniform inputting standards and marking protocols that accurately and uniquely describe and identify each document, the Forum simplifies the work of authors, editors and librarians, and reduces related publishing expense. And the technical expertise exists to begin building it now.

A growing number of researchers and information professionals recognize that scholarly communication is at a crossroads; many are seeking innovative solutions on their own to the wide variety of technical challenges that networked alternatives present. While much visionary work has emerged, the absence of any significantly new prototype for exchanging and preserving research results beyond xxx.lanl.gov suggests the advantages that may accrue from a more broadly-based, collaborative approach.

A Consortium of universities, committed to developing and maintaining an integrated platform supporting all aspects of the scholarly communications process, also provides a basis for conducting meaningful experiments. Universities have the necessary critical mass of participants from varied disciplines. University faculty are already well represented on present editorial boards and include many editors; strong representation of university faculty on the new editorial boards established under the auspices of the Scholar's Forum continues this tradition. Universities have close ties to professional societies, have expertise in information technology, and have a large pool of creative student programmers who can contribute to the infrastructure developments that will be needed. Since universities are responsible for most of the work that appears in the scholarly literature, well-defined, committed administrative support can take advantage of major economies of scale to curtail costs as access to the scholarly literature is enhanced.

Consensus is growing among scholars that change is desirable, and few would not agree that universities possess the talent required to make it so. However, an individual university will have difficulty redefining the paradigm for scholarly communication on its own. The Scholar's Forum provides the basis for universities to apply available talent and resources towards developing practical response to this critical issue.
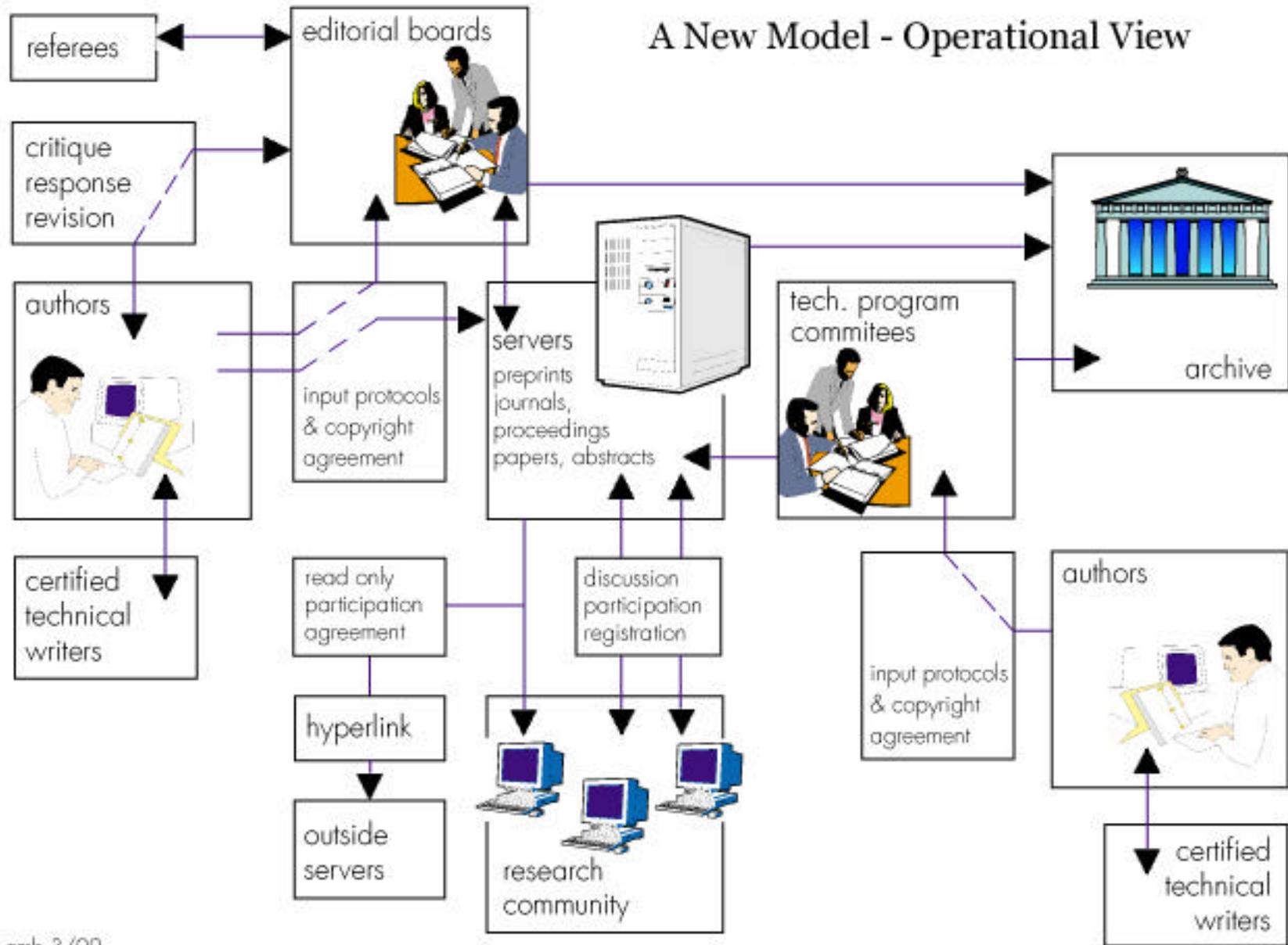
ONGOING DISCUSSION

Posted periodically following Forum text at:
http://library.caltech.edu/publications/ScholarsForum.

A New Model - Operational View

referees

editorial boards

critique
response
revision

authors

certified
technical
writers

input protocols
& copyright
agreement

servers
preprints
journals,
proceedings
papers, abstracts

tech. program
commitees

archive

read only
participation
agreement

discussion
participation
registration

hyperlink

outside
servers

research
community

input protocols
& copyright
agreement

authors

certified
technical
writers

amb 3/99

A Consortium of Universities - Administrative View

copyright management

external fundraising

financial management

consortium of universities

libraries

campus IT org.

standards & protocols

archive

servers

systems

document preparation consultants

editorial board

—— management

➤ maintenance

amb 3/99