

Creating a Sustainable Scholarly Communication System

Kenneth M. King
5/27/04

Abstract

The current process of scholarly publication is widely regarded as unsustainable. Ensuring that scholarly information remains accessible to the world's scholars will require the work of a consortium of major research universities. A global consortium of research universities would have the power to negotiate a mutually beneficial relationship with cooperating publishers including permitting the open publication of preprints in disciplinary archives. This consortium could be built around a shared global electronic library constructed from components managed by individual cooperating institutions. These components built on Open Archives Initiative (OAI) compliant servers using open software (e.g. DSpace and EPrints) are currently installed at many universities. The shared library could look like an extension to an individual member's library and contain a full range of materials certified in a variety of ways by contributing institutions. In addition to publishing books, articles, course materials, videos and databases, universities could individually or cooperatively host open and subscription-funded journals in digital form. They could support open, discipline focused preprint archives and encourage faculty to publish in journals that permitted this. This system would help integrate and coordinate multiple efforts to promote open publication and enable open and not-for-profit publishers and university libraries to become partners in the scholarly enterprise, each responsible for a certain phase of the process.

Faculty Support and Participation is Essential

The first prerequisite of a sustainable system is to provide faculty with incentives to participate. For faculty authors, publication in journals permitting open publication of preprints in disciplinary archives has a number of advantages:

1. Broad and timely dissemination of their work.
2. Greater impact of faculty scholarship as revealed by more citations.
3. Disciplinary and institutional archives to preserve and disseminate their work to future generations; permanence is not provided by storage on personal web pages.
4. Global access through general search engines like Google.
5. A public time-and-date stamp for preprints that contributes to validating ownership of ideas.

In addition, by supporting open and not-for-profit journal publishers, they have an opportunity to help relieve the financial crunch facing their institution due to the rapidly increasing cost of commercial publications. Those costs certainly have an impact on an institutions ability to provide a wide range of scholarly information to its faculty and students. They also have an opportunity to contribute to fixing and reshaping a scholarly communication system that is unsustainable.

Electronic publication also has advantages over publication in print. Authors can include multimedia materials and systems can be implemented that track the evolution of submissions and update them by referencing citations, reviews, and addenda. Also, it provides authors with the opportunity to benefit the global citizenry by providing them with information accessible through search engines that has been certified by an explicit process. Commercial publishers would have difficulty matching the advantages to faculty authors of not-for-profit publishers cooperating with Universities to create a new scholarly information dissemination system.

The Proposal

We propose creating a system that globally facilitates open publication of a wide range of materials that are certified in a variety of appropriate ways using the model of the Internet First University Press (IFUP) at Cornell. This system can be realized as a cooperative venture of universities and colleges around the world to create a global shared library by interconnecting local repositories.

The result would be an enormous electronic library containing the scholarly material that the cooperating institutions, scholarly societies, and disciplines believed was worth sharing, built by coordinated and shared effort. There are recent examples of what can be achieved by harnessing collective university energy. These are Bitnet (Because its time net) and the Internet. Bitnet made possible global scholarly communication at low cost to the institutions participating. It was started by connecting City University of New York and Yale University. Within a few years it connected all the major universities in North America, Europe and Asia – fueled from its inception by the messianic zeal of Ira Fuchs and his disciples who committed volunteer labor to managing it and expanding its capabilities. In order to connect to Bitnet, institutions had to agree to do the following three things:

- a) pay for a telephone line connecting them to a computer port provided by a connected institution,
- b) provide a port to at least one other institution wishing to connect,
- c) pass information coming to it destined for another institution down the chain defined by a routing table.

Thus, for the cost of one communication line, each connected institution had access to global shared bandwidth.

What became the Internet was started when the National Science Foundation paid for a communication backbone connecting the National Supercomputer Centers (NSFnet). With local, state and federal money, universities formed regional consortia to connect their universities to the backbone, sharing bandwidth to reduce communication costs. The same people that were involved in creating Bitnet were largely involved in creating these regional networks. Corporations were allowed to connect to NSFnet subject to the restriction that the information interchange always involve a university at one end and the

exchange be consistent with the educational and scholarly mission of that university. Within a few years NSFnet was connected to similar networks connecting universities all over the world. In 1993 NSFnet was privatized to become the Internet.

A few hundred universities worldwide that initially connected to Bitnet and foreign associated networks proved the value of electronic communication to the remaining institutions and enabled the subsequent creation of NSFnet that led to the Internet. These were primarily the major research universities, the same institutions that produce most of the scholarship published in journals. If those same institutions could be mobilized once again, they could enable a revolution in the way scholarly content is disseminated. The global Internet enables global multimedia communication at low cost. A global network of connected libraries could provide access to all the information assembled by participating universities at little more than the cost of maintaining a local library.

The Bitnet Model

1. It required a vision and a goal that could be achieved by cooperative effort.
2. The benefit resulting to an individual or organization from their effort in participating is proportional to the sum of the efforts of all of the participants.
3. The technology required to implement the goal is available.
4. A person or group of people with messianic zeal is available to articulate the vision and to start the process of implementation by convincing a small number of individuals or organizations to join them in creating a prototype implementation.
5. The idea is contagious and the early implementers recruit and assist new individuals or organizations in joining the effort.
6. An organization is formed to sustain and manage the effort and expand its scope and functionality.

Application of the Bitnet Model to the Creation of a Global Library

The goal is to create a global library of materials worthy of sharing created through cooperative effort and to develop a sustainable model for scholarly information exchange. The benefits of having this library to a participating university would clearly be proportional to the sum of the efforts of all of the participants. The technology required to implement this goal is available. OAIster at the University of Michigan is an example of a library created from OAI compliant components (see <http://oaister.umdl.umich.edu/o/oaister>) and Google currently provides a metadata searchable index into materials stored in a number of DSpace repositories.

Realizing this goal requires messianic leadership and a small number of universities willing to demonstrate its worth to all of the other universities. This effort could result in a global organization with the clout to negotiate a mutually beneficial relationship with not-for-profit journal publishers. This system would enable not-for-profit publishers and university libraries to become partners in the scholarly enterprise, each responsible for a certain phase of the process.

An Implementation Plan

Harvesting metadata from sites that operate OAI compliant servers and creating a common catalog at one site could create a prototype of a global library. There are more than 100 universities with the technology enabling them to participate. Google has already implemented a system enabling federated searching of DSpace repositories.

The cooperating sites could create a membership organization restricted to institutions committed to creating a global library providing content conforming to shared standards and to developing a sustainable model for scholarly information exchange. Their first task would be to develop a common set of goals. Possible goals include:

1. to create a global federated library of open materials that the cooperating institutions believed were worthy of sharing and that conformed to agreed upon standards of certification and presentation,
2. to encourage faculty to submit preprints of their scholarship to an existing preprint server (e.g., Arxiv) or to an equivalent disciplinary archive and to publish in Journals that permit this,
3. to support open journals and journals that permit open publication of preprints in disciplinary archives (see <http://www.ecs.soton.ac.uk/~harnad/Temp/Romeo/romeo.html>), and
4. to create a global organization to coordinate efforts; to support the growth, maintenance and management of the global library; to enhance the software needed to facilitate global access to this distributed library; and to create a structure for the global library that facilitates the growth of discipline communities within it.

A structure that facilitates the growth of discipline communities within the global library could be created by linking discipline communities in local repositories to create a global discipline based community. An alternative path would be for members of a discipline group to link their departmental servers running DSpace or other OAI compliant software in a way that created a disciplinary component of a global library.

Success at a small number of universities in creating a common electronic library could lead to export of the system to other universities. For institutions running DSpace, EPrints, or an equivalent archiving system that is OAI compliant, the incremental cost of creating and managing a local component of a global library should be small. Thus a group of universities could build a global library out of a distributed network of local university servers with a modest investment in support of faculty contributing to their local OAI compliant repository. The content of the full universe of materials at all the participating institutions would look to readers as though it were a single unified library collection presented as an extension to a member institution's library.

Benefits of Participation

The benefit to an institution of participation in building a global library could include the following:

1. Ability to participate with other elite institutions in defining and building the library of the future.
2. Access to a huge library of materials that they did not have to store, preserve or process.
3. Ability to showcase their institutions' scholarship including a catalog of information available for sale, subscription, or pay-per-view. Inclusion of information in a global library would have far more impact than inclusion in a local repository.
4. Ability to benefit from economies of scale and elimination of redundant effort through consortia cooperation. For example, economies of scale can be achieved in archiving, dissemination, and in preserving material as technology changes. Redundancies in storing and processing information could be eliminated.
5. Having their content in a Google index and available under a search button covering members' contributions.
6. Access to information about their authors' reviews, downloads and citations.
7. Access to other software and services cooperatively developed.

The Cost of Participation

There are two major cost components, the cost of maintaining the hardware and software system supporting an archive and the cost of developing content. The cost of maintaining the DSpace server at Cornell is approximately \$100,000 per year. Component costs include hardware amortization, software system installation and maintenance and content backup. It is easily possible for a number of institutions to share a server.

The cost of developing content and support for faculty contributors depends on the volume and type of content created. Adding theses and preprints is largely an author responsibility and inexpensive but requires some user support. The same is true of other author-contributed material created in digital formats. Scanning grey scale material costs of the order of 25 cents per page and this material can be run through optical character recognition software for a few cents more per page. Creating complex documents and multi media materials is significantly more expensive. Volunteer labor has greatly reduced the cost of creating the content in the Internet First University Press at Cornell. However, we have discovered that faculty encouragement and support is essential to getting high quality content. The Cornell experience in creating a digital archive is discussed in more detail in Appendix A.

Extending the Scope of the System

Some additional software would be needed to enhance the utility of the global library, particularly if it wishes to support and host journals and books created by groups of scholars that publish materials that go through processes similar to those of the Public Library of Science (PLOS) and university presses. Most of that software already exists but some of it is not open software. Key components of a full function local component library are:

- a. OAI archive software (like EPrints and DSpace),

- b. software supporting the end-to-end processes of journal and monograph creation and dissemination (like Open Journal Systems, an open source program produced and currently supported by the Public Knowledge Project at the University of British Columbia),
- c. desktop publishing software enabling authors to adhere to submission standards (such as Adobe Acrobat, EndNote and LaTeX)
- d. a catalog identifying the contents of the global library and access pointers,
- e. a preprint distribution system (like ArXiv)
- f. linkage to a search engine that supports full text indexed access to any user with a web browser,
- g. software for tracking certification status, reviews, citations and other information relating to the historical evolution of the information, and for updating submissions to reflect these changes, and
- h. a user-friendly interface that simplifies the process of search and submission.

A drag and drop icon on the desktop of faculty could simplify submission and encourage participation. It could also provide faculty with a list of cooperating journals.

A distributed system implementation plan would enable collections of institutions or disciplinary groups to operate a single server at shared cost or one or more disciplines sharing a preprint server such as ArXiv. It would enable groups with a shared interest such as disciplinary groups to connect local servers in a way that created a sub library in the global library serving their interests. The global library could interconnect to services that support subscription and pay per view access for materials where open access was not currently financially feasible and other services such as print on demand. There are materials for which high selectivity and expensive publication processes or the necessity for restricted access could make producer-side funding of open publication hard to afford or impossible to implement. However a global library organization could enable and encourage a transition from high priced commercial journals to lower cost open journals with high selectivity and publication standards like the PLoS.

Effect on Stakeholders

Major stakeholders in the current system are

- a. readers, consumers of information,
- b. authors, producers of information,
- c. libraries, collectors and disseminators
- d. open publishers, packagers and marketers
- e. not-for-profit publishers

- f. commercial publishers, and
- g. universities

Consumers of information want rapid and low cost access to the information they need when they need it. They value services that organize and certify information in a way that saves them time and provides access to information needed for a task at hand.

Scholars who produce information are interested in advancing their discipline, enhancing their careers and reputation, the widest possible dissemination of their ideas, and in benefiting financially from the information that they produce if it has economic value. In the current system, scholars receive the highest reward by publication in journals with distinguished editors that recognize the elite members of their discipline. Increasingly, they also value broader and timelier dissemination of their ideas, finding frequent references to themselves and access to their scholarship in Google, increased citation counts, and the promise of institutional archives to preserve their work for future generations and to disseminate it globally. Also, an ArXiv like preprint service registers ownership of ideas. The posting has a time and date stamp. Theft of ideas is more difficult than in the traditional decentralized system that has no such system for staking an intellectual claim. In addition, by supporting not-for-profit journal publishers, scholars have an opportunity to help relieve the financial crunch facing their institution due to the rapidly increasing cost of commercial publications and to help fix and reshape a scholarly communication system that is unsustainable. Finally, they have the opportunity to benefit the global citizenry by providing them with information accessible through search engines that has been certified by an explicit process.

Libraries traditionally support access to information as a right of every citizen or at least every member of the community that they serve. Libraries are for the most part funded in a manner that enables them to provide information to their clientele at no charge. The amount of information that they can provide is constrained by the resources available to them. The high cost of commercial journals is a significant factor in reducing the information that they can provide. In the global library system, libraries are the publishers supported by highly automated tools to archive, organize, and retrieve materials. The emerging electronic technology will allow libraries to avoid much of the handwork of housing, processing, and circulating materials. Some of the savings accruing from these cost reductions could be used to support operating the local library and faculty contributing to it. For libraries already operating an archive, incremental costs should be small. However an indispensable part of this effort requires proactive support of contributors to the library and one place to put that support is in the library. Just as the invention of the moveable type printing technology by Gutenberg freed monks from their transcriptional responsibilities, reduction in the responsibility to process and store information can free librarians to focus on supporting and managing a global library and more sophisticated collection development, search, dissemination and retrieval activities. It would also enable them to take a more active role in assembling and disseminating their institution's scholarship as a key member in the scholarly

information chain representing the interests of the public and their institutional clientele.

With some oversimplification, commercial publishers owe their primary loyalty to shareholders and seek to control access to information in a way that maximizes their profit. They are often attractive to authors because they can afford to recruit an elite group of paid editors and because they don't charge submitters page charges for submissions. They generally do not accept pre-published articles and insist on owning the copyright to the material that they publish. The creation of a global library might retard the growth of commercial journals because global access to preprints at no charge with perpetual access to back issues in the fashion of JSTOR may be more difficult for commercial publishers to achieve.

Not-for-profit publishers seek to advance the disciplines they serve and their members by providing a communication channel for information and various member services at the lowest possible cost consistent with maintaining financial viability. They will resist any new system that threatens their financial viability. Finding a way to encourage their participation will be a major challenge for this initiative. As information is increasingly available to scholars through institutional electronic subscriptions, scholars are dropping individual subscriptions and relying on institutional subscriptions. Thus institutional subscriptions are increasingly important to not-for-profit publishers. This gives a large coalition of libraries a special opportunity to work with not-for-profit publishers to advance both groups' interests. Establishing a symbiotic relationship with university authors and libraries will significantly increase their volume of contributions. If the cooperating institutions could reduce their dependence on high priced journals, they could afford to pay more to cooperating not-for-profit publishers who would be publishing a larger volume of articles. A global consortium of research universities would have the power to negotiate a mutually beneficial relationship with cooperating publishers. This for example could include longer-term bundling arrangements similar to those used by commercial publishers. This system would enable not-for-profit publishers and university libraries to become partners in the scholarly enterprise, each responsible for a certain phase of the process. This partnership could fix a system that is "incontrovertibly unsustainable."

Important Next Steps

A small group of universities should create a demonstration of what can be achieved by linking their repositories and agreeing on interim membership standards, policies and goals. They then need to export the idea of participation to other universities by demonstrating benefits of participation.

This group should reach out in an effort to get advice and broad support from key stakeholders. Meetings with these stakeholders (faculty, librarians and not-for-profit journal publishers) could inform the definition of the goals, policies, and standards needed to launch the Global Library.

A cooperative relationship with not-for-profit societies that ensures their financial viability as this global library evolves is essential. A good starting point would be to identify a few journals whose editors would be willing to work with libraries to define the terms of a mutually beneficial relationship that could be generalized. Existing open access journals should be willing to participate to achieve greater visibility. Editors of new open and subscription journals might choose the Global Library to achieve economies in production and full integration with a global digital virtual library.

A critical next step will involve encouraging faculty to publish in journals permitting open publication of preprints and to deposit preprints of their articles in an open disciplinary archive. An effort should be made at institutions participating in this project to secure the support of faculty leaders at their institutions. If it can be shown that cooperating journals thrive, other journals will be willing to participate. Widespread faculty commitment to this new scholarly dissemination system would drive the whole effort and insure its success.

Malcolm Getz's white paper has demonstrated that research institutions could afford large scale producer-side funding of open publication of journal scholarship in part through savings associated with the reduced storage and processing charges associated with electronic publication and in part from being able to reduce the average cost of article publication to that associated with not-for-profit journal publishers. A key first objective therefore should be to encourage scholars to replace high cost journals with not-for-profit journals with lower publication costs. A global library would support that effort by enabling collections of faculty to create new journals at low cost.

Success in demonstrating progress could lead to early endorsement by key presidential, library and other organizations with an interest in fixing the current scholarly communication system. An important symbolic step would be for presidents of universities, library associations, and university faculty organizations to endorse the open publication of preprints of articles and to encourage faculty to submit their articles to journals that permit this practice and access to back issues through a service that takes long-term preservation seriously.

We have an opportunity to transform in a fundamental way how scholars record and share their best works for the planet and for the future.

APPENDIX A

The Cornell Model of a Local Component of a Global Library

At Cornell we have built an Internet First University Press (IFUP) see:
<http://dspace.library.cornell.edu/handle/1813/62>.

The IFUP at Cornell permits us to publish books, articles and video materials at low cost. It has been implemented as a special community within DSpace designed as a prototype of how an institutional component of a global library

might look. Low cost has been achieved through (a) the use of open technology for publication, DSpace, (b) desktop publishing software, and (c) by maximizing the impact of volunteer labor by decoupling all of the elements of publication from certification to dissemination. Further cost saving should come from economies of scale, particularly those arising from cooperation with other institutions in creating a global library. An indispensable component of this effort has been proactive support and encouragement of the people contributing to the IFUP. Without that support, very little would have happened.

The initial materials in the Cornell IFUP have been certified in a variety of ways. Some have been certified through a peer review process that is essential for certain categories of publication. Some have been certified by the process that created them and the reputation of the people participating in that process and some by the reputation of the contributors. The Cornell Graduate School has implemented procedures to enable any of the approximately 2,000 students graduating annually to deposit their theses in the IFUP. Hans Bethe, a Nobel laureate with the help of Cornell colleague, President Emeritus Dale Corson, created a series of videos. A book edited by Jay Orear contains contributions from distinguished scientists about Enrico Fermi. Jack Oliver who contributed two books is a distinguished geologist who was a major contributor to developing the theory of plate tectonics. He has contributed his previously published memoirs and a book that was published earlier but is now out of print. Richard Rand has contributed notes from a highly successful course on nonlinear vibration theory developed and taught over many years at Cornell. The Cornell Cooperative Extension service has contributed articles on a number of subjects. These articles are certified by the reputation of the service and the processes that it uses to insure that its advice conforms to the best science available. We are in the process of putting up high resolution images of plant cells whose pedagogical and research value is certified by its technical parameters and its enthusiastic users.

IFUP authors forego royalty income for the materials distributed online, but do receive a royalty for printed copies obtained through our print-on-demand service. Unlike traditional publishing, the authors retain copyright ownership and the right to re-purpose the content without the necessity of seeking permissions, (as often must be done even for their own classroom usage). IFUP authors grant only a non-exclusive license to distribute their materials.

The current contents of the Cornell IFUP can be found at:
<http://dspace.library.cornell.edu/handle/1813/62>.

Authors or other volunteers contributed most of the effort required to get these materials into the IFUP. In some cases elements of the process like audio or video editing, and printing were purchased from university services. In some cases, manuscript editing was purchased from freelance editors. Our experience has been that faculty possess a range of materials that with a little help they are delighted to place in an archive that will be Google indexed and shared with a global audience. This provides a new lease on life for materials such as out of print books. They also appreciate the fact that the material will be preserved by a system that will make it accessible to future generations. The widespread current

practice of storing articles on personal or departmental servers does not insure long-term preservation.

An ad hoc review committee (the DSpace steering committee at Cornell) has approved the inclusion of materials assembled to date based on the criteria:

- a. Does the material have scholarly merit?
- b. Does the material have probable value to the global scholarly community?

We are in the process of creating faculty review boards responsible for approving future contributions. We believe that the certification processes described above are appropriate for a broad class of materials like those currently in the IFUP at Cornell and for preprints and post prints of journal articles that are disseminated in services like ArXiv. However, articles in journals need to be certified by much more formal processes involving external peer review. The Cornell IFUP is prepared to host open journals with a quality peer review process. Universities participating in the operation of a global electronic library will need to establish internal review procedures that protect against the distribution of academic spam.

Important Elements of a Low-cost System.

The opening paragraph of the Public Library of Science (PLoS) white paper on Publishing Open-Access Journals states:

"There are many different paths to producing a journal, either online or in print, with a tremendously wide spectrum of costs that can be generated or avoided during the publishing process. The aggregate cost of shepherding manuscripts through peer review, preparing selected papers for publishing, and finally disseminating articles depends on the particular steps that a publisher deems necessary for a particular journal. Using unpaid academic editors and an open source online journal management system, eschewing frills in the production process, and publishing online directly in archives with minimal formatting requirements (for example, those that accept articles as simple PDF files), a publisher could potentially produce a peer-reviewed journal spending little or no money. Additional steps in the publishing process—from copyediting, to using professional editors, to formatting articles to meet the more stringent technical standards for online publishing and for archiving in repositories such as PubMed Central—can increase costs marginally or substantially, depending on which components of the process are added. For PLoS Biology, the Public Library of Science [PLoS] chooses to include many steps beyond the minimal requirements for overseeing peer review of submissions and for online hosting for accepted articles; consequently, our costs are considerably higher than those that other predominantly electronic journals might generate in the publishing process."

We believe that the Cornell experience in publishing low cost materials can be generalized to a global system for collectively disseminating materials of scholarly value. Where particular intellectual purposes require it, IFUP

publications can add dimensions at extra cost along the lines of PLoS. A major goal of our project is to encourage open dissemination of preprints and to facilitate and support open journals.

An important element in lowering cost is decoupling the processes required to produce and distribute knowledge and enforcing standards that define a limited number of formats for documents submitted. The PLoS white paper on Publishing Open-Access Journals details the steps and costs in the publishing of their journals and they have outsourced many of them. Their costs for each step are probably close to the top end of the scale because of the complexity of the material they publish as well as the quality of the end product they wish to achieve. For example their manuscript editing costs are \$20 per page. Some book presses hire experienced freelance manuscript editors for \$3 to \$6 per page depending on the scale of the effort required. It is clearly possible to publish material like preprints, and proceedings of conferences where immediacy is the highest priority and materials like out of print books, and faculty archives of previously published material at very low cost as we have demonstrated with the IFUP at Cornell.

Perhaps the best example of what can be achieved by decoupling certification from publication is the ArXiv system developed by Paul Ginsparg. ArXiv is an openly accessible, moderated repository for scholarly papers in specific scientific disciplines. ArXiv publishes articles for a cost of less than \$10 per article. This system has had a catalytic impact on Physics where the immediate availability of new ideas and discoveries is often more important than waiting for materials to go through a careful review and certification process. ArXiv, however, accommodates both preprints and published articles. It has recently introduced a system requiring new submitters to be endorsed by a previous submitter.

By examining the 2003 ISI index, Phil Davis, a Cornell Librarian, has found 5465 occurrences of the name of a Cornell author in an article in a journal indexed by ISI. Correcting for publication in journals not indexed by ISI, Phil estimates that a Cornell author's name occurred in about 6000 articles published and they were lead authors in about 3600 cases. Electronic publication of preprints of every Cornell authored article would probably cost less than \$36,000 multiplied by the ratio of submitted articles to articles accepted for publication if the lead author paid, and if we had a distribution system with volumes and costs comparable to ArXiv.

There are a number of disciplines where books are of interest to an audience too narrow to permit breakeven publication by a commercial press. We have published books in the IFUP at Cornell for a few thousand dollars compared to costs more than 10 times greater for traditional book publication. This suggests that if scholars in these disciplines were to create a low cost certification system, electronic publication with some institutional support could enable dissemination of their scholarship. Most books published in IFUP will be fully word-searchable by Google on the Internet and embedded in the worldwide IFUP database. (This capability is expected to be provided in an early DSpace release.) In addition, authors could receive royalties through a print-on-demand system that the IFUP supports.

There are authors willing to forgo the royalties that they might receive from commercial publication in return for wide exposure and the careful reading and review of their work by leading scholars associated with open publication. We have published examples of these. As the PLoS article points out, it is possible for scholars willing to create journals with volunteer labor to publish openly at very low cost. There are many examples of these.