



## Online publishing: The arguments in favor of free access

by Dan Wise

Taking advantage of the broad range of opportunities offered by the Web is a challenge that excites many scientists, along with a host of others. The argument presented is that information wants to be free. However, Tim O'Reilly, as president of O'Reilly & Associates, writing as an interested observer, states, "Information doesn't want to be free. *Information wants to be valuable.*" The debate now is how to use the vast opportunities afforded by the Web to make enormous quantities of scientific information, data, and opinion freely available and valuable to the widest possible audience.

The proposal that sparked the current debate was a demand by the Public Library of Science (PLS) that all publishers of scientific journals place their total journal content in a central online repository managed by PLS within six months of publication or face a boycott by scientists worldwide that would strip the journals of both authors and peer reviewers. As of mid-August 2001, more than 26 278 scientists from 170 countries have signed the petition and pledged their support of the boycott.

In this article, I've summarized the views of the several vitally interested parties in this debate, all of whom have a stake in making sure that whatever decision is made serves the best interest of their sectors of the information transmittal and retrieval process.

### Views of scientists

Given that a scientist's reputation and prospects for future growth in both stature in the field and in financial position depend on exposure within the scientific community, the

---

### Contents

Special issue: Online publishing of science	
Online publishing: The arguments in favor of free access	1
Editorial: Publishing or perishing in an online world .....	2
Online publishing: Scientific communication on speed ...	5
Join us on the Science SIG mailing list .....	8

---

argument for free access to all published materials is obvious. Every scientist wants the widest possible dissemination of every word they have published. Indeed, this is one of the major arguments proposed by numerous scientists in support of the boycott. There are, however, others who see the timetables proposed by PLS as being unrealistic.

Understanding also that timely access to published information may be vital to the progress of scientists' current projects, the argument for free access is also obvious. Michael Eisen and Pat Brown, writing on behalf of the PLS initiative sum up their argument this way:

"Scientific progress and public welfare would be much better served by a scientific literature that belongs to the public, accessible and usable by anyone, anywhere, without barriers, charges or restrictions. Private ownership and monopoly control of the scientific literature blocks the free flow of scientific knowledge. It prevents independent creative scientists from exploring and developing new ways to integrate and organize this rich but sprawling and fragmented body of knowledge. It is unfair to scientists. It is unfair to the citizens of the world, who have paid for most of the research but are denied its full benefits. And there is absolutely no evidence that private monopoly ownership is the only practical business model."

*"Every scientist wants the widest possible dissemination of every word they have published."*

### Views of librarians

Librarians in general are facing serious pressures on acquisition budgets. Ann Okerson,

(continued on page 3)

## Editorial: Publishing or perishing in an online world

By Geoff Hart (with input  
from Marilyn Barrett-  
O'Leary, Laurel Busch, and Dan Wise)



“...should you let authors include the citations, with full knowledge that much of the information will soon become unavailable to readers?”

In April 2001, *Nature* launched an online debate on a hot topic in scientific publishing: the impact of online publishing of original research. Stimulated by a challenge to create immediate, unrestricted access to the primary literature, the debate became a spirited exchange. Whatever the final outcome of the debate, we'll unquestionably see more scientific information moving online as the Web becomes an increasingly important part of our personal and professional lives. In this issue, Dan Wise and Marilyn Barrett-O'Leary share the lengthy task of reviewing the online discussion and summarizing both sides of the debate. I hope their hard work will stimulate a fruitful discussion among Science SIG members, whether here in the newsletter, in our online discussion group (see page 8 for details), or in both forums.

I find this topic particularly interesting in light of the impermanence of online information. Learning how to cite online information is easy enough, since many guidelines now exist (e.g., Thibault 1998), but *whether* to cite them poses a far more difficult problem: If the purpose of reference lists and

bibliographies is to help readers find original sources of information, how do readers cope when the online information disappears? Laurel Busch reported that about half the URLs cited in the journal articles she edits no longer exist by the time the articles are accepted for publication. Davis and Cohen (2001) confirm this finding: in student papers written between 1995 and 1999, only 18% of the URLs in papers published in 1996 still led to the correct online documents, versus only 55% for papers published in 1999.

As an editor, should you let authors include the citations, with full knowledge that much of the information will soon become unavailable to readers, or reject citations of online reference material? Treating such material as personal communications certainly reflects their transient nature, but sidesteps the problem. Instead, I propose something more radical: create a precedent by insisting on a practice parallel to that used for PhD and other dissertations, in which the sponsoring university archives the materials in their library. For example, a university press, aided by a skilled librarian, could insist that any Web-based material cited by an author be provided to the university library for archiving—probably in print, initially, but as libraries evolve into places where electronic information achieves status equal to that of printed information, a CD or other digital archival copy should also be required. Subsequent literature citations would include

(continued on page 5)

*the Exchange*, Vol. 8 No. 3, October 2001.

*The Exchange* is published on behalf of the Scientific Communication special interest group of the Society for Technical Communication. Material in *the Exchange* can be reprinted without permission if credit is given to the author and a copy of the reprint is sent to the editor. Please send comments, letters, and articles to the editor.

**Editor and publisher:** Geoff Hart, c/o FERIC, 580 boul. St-Jean, Pointe-Claire, Québec, H9R 3J9 Canada  
(geoff-h@mtl.feric.ca)

**SIG manager:** Amy Burdan (BURDAN\_AMY@LILLY.COM)

© 2001, Society for Technical Communication (901 North Stuart St., Suite 904, Arlington, Virginia 22203-1822 U.S.A., 703-522-4114, 703-522-2075 fax, www.stc.org)

*“Free access...” (continued from page 1)*

associate university librarian at Yale University, presented data from 103 libraries showing that expenditures for serial acquisitions had doubled from 1986 to 1999. Okerson states that even the wealthiest institutions have budget constraints that prohibit them from acquiring more than a fraction of the more than 20 000 scholarly journals currently available. Her observations in writing for this on-line debate support the views of the scientists:

“Scientists demand the broadest possible exposure for their own work and access to the work of others, particularly through the deployment of new information-dissemination technologies. The library marketplace that pays for journal subscriptions likewise argues for the broadest access at the most affordable price.”

She goes on to acknowledge that “few would argue that the required information is failing to reach those who are able to contribute if they have the necessary information.” Okerson represents a widely held view that many of the most prestigious journals, mostly published by large for-profit organizations, are overpriced. Lundy H. Pentz, associate professor of biology at Mary Baldwin College, commented that their library budget permits acquisition of very little beyond *Science* and *Nature*. Richard Luce, director of the research library of Los Alamos National Laboratory, observes that “... publishers, secondary providers, aggregators, and libraries are already under siege, and market forces will drive substantial change in the roles they currently play.”

Addressing the concept proposed by PLS, Luce notes, “The notion of a centralized archive has some attractive advantages, notably the convenience of a one-stop shop and presumably serving as the enabling mechanism to access standardized metadata.” He also acknowledges the arguments of others that “the efficiency of large-scale searching on a single site, extensive citation interlinking between reports originally published in diverse journals, and linking to

other types of data” are advantages of such centralization.

#### **For-profit publishers**

Only two for-profit publishers chose to comment in this debate: Elsevier Science and Blackwell Science. As one would expect, most of their comments were negative. However, Derk Haank, CEO of Elsevier, acknowledges that “Boundary-free access from ‘past to current’ and ‘current to past’ will challenge publishers to explore new ways of developing new services for scientific communities, and serve them better.” Haank believes “the future role for publishers lies in developing these new products and services, of which content forms an integral part.” Haank said that Elsevier aims to give scientists desktop access to all information they need *at a reasonable price*.

#### **Not-for-profit publishers**

Not-for-profit publishers tended to be either alarmist (the smaller journals and those published by smaller technical societies will have to cease publication) or pessimistic (this proposal just isn’t workable for a variety of reasons). However, Martin Richardson, journals publishing director for Oxford University Press, cited current initiatives of his organization to make journal content available 12 to 24 months after print publication. Richardson cited statistics showing that the number of times *The EMBO Journal* was accessed per month dropped off dramatically only 3 to 4 months after publication. He states that there was no detectable increase in usage once the journal was made available online for free through HighWire Press.

Although Richardson feels that archives should be available on a fee-paid basis, he acknowledges that “Online publication has greatly increased the dissemination of the journal literature, allowing easy access from the office, laboratory, home, or airport. Unlike their printed counterparts, online journals are

*“...even the wealthiest institutions have budget constraints that prohibit them from acquiring more than a fraction of the more than 20 000 scholarly journals currently available”*

*“... many of the most prestigious journals, mostly published by large for-profit organizations, are overpriced.”*

*(continued on page 4)*

*“Free access...” (continued from page 3)*

*“...at least some authors are willing to pay for instant access and... ‘reasonable’ fees ought to be charged by journals.”*

*“... the physics ‘eprint archive’... archives approximately 30 000 new articles annually, a number that increases by 3500 each year.”*

available 24 hours a day, 365 days a year. They are often fully searchable and are linked to other journals and books, and bibliographic and other databases. They thereby provide a wide range of services for the scientific communities they are designed to serve.”

Thomas J. Walker, of the University of Florida, points out that at least some authors are willing to pay for instant access and suggests that “reasonable” fees ought to be charged by journals. Martin Blume, editor-in-chief of the American Physical Society, also advocates a fee system to cover the costs of peer review and ensuring quality content.

### **Databases and repositories**

Perhaps one of the most telling arguments for free access was made by Stevan Harnad, of the University of Southampton, in addressing the self-archiving initiative:

“There are currently at least 20 000 refereed journals across all fields of scholarship, publishing more than 2 million refereed articles each year. The amount collectively paid by those of the world’s institutions which can afford the tolls for just one of those refereed papers averages \$2000 per paper. In exchange for that fee, that particular paper is accessible to readers at those, and only those, paying institutions ... [N]ot even Harvard can afford access to anywhere near all of the literature. Hence, most refereed articles are inaccessible to most researchers ... From the authors’ viewpoint, toll-gating access to their findings is as counter-productive as toll-gating access to commercial advertisements.”

Walter Warnick, director of the Office of Scientific and Technical Information (OSTI) of the U.S. Department of Energy, notes that although it is technologically feasible to build a system that would allow students and researchers to conduct literature reviews on the Internet, such a system does not currently exist. He comments that there are some 7000 scientific

and technical preprint sites, but most of them lack formal data structures and they are not linked. The advantage of such preprint sites, according to Stevan Harnad, is that the manuscripts of papers appear up to 12 months before they appear in refereed journals. Harnad cites an example of the popularity of preprint sites, the physics “eprint archive”, which archives approximately 30 000 new articles annually, a number that increases by 3500 each year. This archive has 14 “mirror” sites worldwide and gets about 160 000 user hits each weekday at its U.S. site alone. His conclusion is that self-archiving is feasible and that such archives are heavily used.

Librarian Richard Luce also mentioned the Open Archive Initiative, which was created to solve inoperability issues between self-archiving sites by establishing standards and protocols that ensure the archives work together.

### **Reader feedback**

Joseph Ryan, of the National Human Genome Research Institute, notes a further benefit of having centralized open access. In addition to being able to conduct full-text searches and then retrieve articles of interest to read, researchers would also be able to mine the data for meaningful patterns, something that cannot be done readily with self-archiving.

### **A brief summation**

Richard K. Johnson, enterprise director for the Scholarly Publishing and Academic Resources Coalition (SPARC), sums up the issues very well when he says:

“It is important to remember who are the consumers in scientific communication: the authors, who want their work to be widely disseminated and recognized; readers, who would like convenient, barrier-free access; and institutions, which require cost-effective means both to support and evaluate employees, and to teach students.”

Richard Luce observed, in a similar vein, “It would be in our best long-term interests to

*(continued on page 5)*

*“Free access...” (continued from page 4)*

optimize our communication systems to support a variety of approaches while we evolve our understanding of the coming adaptive Web and its impact on the communication of science.”

Johnson also warns that there may be no single solution to broadening access to scientific information. Use of the Web as a vehicle will require experimentation and innovation, with the success of reforms being based on feedback from authors, readers, and institutions. Ω

*“Technology is a queer thing. It brings you great gifts with one hand and stabs you in the back with the other.”—C.P. Snow*

## **Online publishing: Scientific communication on speed**

*By Marilyn Barrett-O’Leary*

By urging accelerated acceptance of open and free online access to science literature and databases, PubMed Central and the Public Library of Science (PLS) are putting scientific communication on speed. Although most interested stakeholders, from scientists to librarians, endorse use of the Internet for distributing scientific literature, they are against rapid change that will disrupt the socioeconomic and academic systems associated with the scientific research process.

For more than a hundred years, science has advanced through literature review, the scientific method, and the peer-review process. The resulting disciplines, based upon evolving bodies of knowledge, provide a basis for university instruction, organization, and tenure, as well as for continuing scientific study. Professional societies, spawned by this process, expedite scientific communication and archive resources.

*“Editorial” (continued from page 2)*

the standard author/date/title/publisher information, plus an indication of where the material is archived so researchers can obtain copies by e-mail or via interlibrary loan.

Jon Shear, a local filmmaker was quoted as saying, “I vowed [that] if I complained about things more than three times, I had to do something about it.” Maybe it’s time we took Jon’s advice. Ω

### **Resources**

- Davis, P.M.; Cohen, S. 2001. The effect of the Web on undergraduate citation behavior 1996–1999. *J. Amer. Soc. Information Sci. and Tech.* 52(4). (Summarized online at [www.people.cornell.edu/pages/pmd8/](http://www.people.cornell.edu/pages/pmd8/))
- Thibault, D. 1998. *Bibliographic style manual*. Section 3. Electronic documents. National Library of Canada, Ottawa, Ont. 7 p.

Caution should accompany attempts to change this system.

This sentiment is reflected in the views of several vitally interested parties. As Ann Okerson, associate university librarian at Yale University says, “The idealism of the moment needs to express itself in a way commensurate with its own principles, by establishing dialogue, building community, and giving standards and consensus time to develop.” The reasons to change slowly are described below.

### **Information exchange and dissemination**

Broader, timely information exchange is already occurring within individual disciplines and groups of disciplines. The Entomological Society of America has integrated four journals online, according to University of Florida entomologist Thomas J. Walker; the American Physical Society publishes three, according to the society’s Editor-in-Chief, Martin Blume. The American Society for Biochemistry and Molecular Biology (ASBMB) is currently publishing online in partnership with Stanford University’s HighWire Press, according to

*“The idealism of the moment needs to express itself in a way commensurate with its own principles...”*

*(continued on page 6)*

“scientists  
... are  
anxious to  
combine  
information  
across all  
scientific  
disciplines  
... [but]  
acknowl-  
edge the  
existence  
of techno-  
logical  
barriers...”

“They  
want ... a  
generally  
accepted  
quality-  
control  
system to  
avoid  
incorpora-  
tion of  
'rubbish'.”

Robert D. Wells, president of ASBMB and Herbert Tabor, editor of the society's print journal. PubMed (one of the organizations that stimulated the debate) is publishing medical research from many sources. Several scientific journal publishers and for-profit publishing groups such as The PrePRINT network and the e-Print Archive at Los Alamos publish online peer-reviewed research articles from several disciplines, according to Walter Warnick, director of the Office of Scientific and Technical Information in the U.S. Department of Energy.

While scientists like Mark Gerstein and Jochen Junker are anxious to combine information across *all* scientific disciplines in order to integrate data and possibly discover heretofore unrealized relationships, they acknowledge the existence of technological barriers to rapid integration, such as the limitations of today's keyword-based search engines. “We expect that complex scientific data sets will become tightly integrated and entwined with the literature, with the interface to publications moving away from simple keyword search models to one reflecting the structure of biological information itself... One might fly through a large three-dimensional molecular structure... where various surface patches would be linked to publications describing associated chemical bonding.” Even this imaginative linkage between full text and archived refereed articles is limited by Gerstein's and Junker's primary interests—their academic disciplines (molecular biophysics, biochemistry and computer science).

Some, like Joseph Ryan of the National Human Genome Research Institute, envision mining the data from multiple sources for meaningful patterns. Warnick and others responsible for databases and repositories, as well as Los Alamos librarian Richard Luce, point out that integrating huge bodies of materials from different disciplines, just like integrating huge data sets from different experiments, requires

planning and process adjustment. Apples and oranges can't be easily compared or integrated.

### **Responsibility for archiving**

Such new databases expand archive function and management. Controls will be needed for responsibly extracting and reusing database information, according to Les Grivel of E-Biosci, an online database operation. His organization is developing a control method in certain disciplines.

The sheer mass of data (some 3 000 000 new articles per year) presents database management and maintenance challenges. Two people involved in the evolving GenBank database, Jo McEntyre and David Lipman of the National Center for Biotechnology Information, describe the need to develop new and innovative software containing access protocols. They want process changes such as agreement on common format and markup language for the entered data, and a generally accepted quality-control system to avoid incorporation of “rubbish”.

With time, these changes can be made, but PLS is demanding immediate compliance from scientists and authors. How will PLS cope with the sudden influx of material into a new and different type of database that can support queries from people in many disciplines?

### **Peer review**

The expertise that has evolved in each discipline, based upon peer review of research and conclusions, is reflected in society membership, honors, and publication in refereed print journals. Once this process is broadened across disciplines, the rigor of review may be diluted by loss of authors and reviewers, says Ira Mellman, editor of *The Journal of Cell Biology*.

Several journal publishers point out that replacing print journals with online journals could even eliminate the peer review process—“... the equivalent of throwing out the baby

*(continued on page 7)*

*“Scientific communication...” (continued from page 6)*

with the bath water”, says for-profit publisher Derk Haank of Elsevier Science. Scientists need the verification provided by peer review, says scientist Steven Harnad from the University of Southampton’s Department of Electronics and Computer Science.

#### **Loss of revenues**

Perhaps the most often-repeated argument against rapid adoption of open online access to primary literature and development of broad, cross-disciplinary archives is the negative impact on revenues that support the current system of scientific communication. Historically, receipts from publication, archive, and subscription fees support the peer-review process, says for-profit publisher Robert Campbell of Blackwell Science, Ltd., and pay print costs, say for-profit and not-for-profit publishers. Some for-profit publishers now use these funds to digitize and publish earlier issues electronically. Some scientific societies also use those funds to reduce the size of conference fees and dues. Thus, free online access might have the opposite-from-intended effect by pricing some scientists out of the discourse, says Martin Richardson of Oxford University Press.

The costs associated with new software, more hardware, and regular online maintenance are currently unclear. In the short term, how will PLS secure resources to fund development of new software and cover the ongoing expenses of database management? In the long and short terms, new funding resources such as government agencies or social organizations will present new problems, since their grants may be based on goals other than pure science, cautions Grivel. Luce speculates that open but not free access may be the solution.

#### **Integrity of online information**

Online open access could mix “gray” literature with primary literature or careful research with careless work. As the publisher of

High Wire Press, Michael Keller, says, “If PMC and PLS were to succeed in their aims, they could weaken the competitive position of scholarly societies vis-à-vis irresponsible publishers.” Owners of the new Web servers housing primary literature may not preserve the author’s original intent, cautions Mellman.

The Internet is not yet secure enough to guarantee the integrity of all data posted online. Today, software, hardware, and human errors, as well as dishonest intent, can jeopardize it. Guards have not been developed to block hackers, who are always challenged by new online products.

As well, precisely how copyright law will apply to this new type of publication remains unknown; recent lawsuits such as the one against the Napster music-exchange service suggest that online doesn’t mean free, but the jurisprudence is still being established. While some regard this as good news—because copyright, retained by journals, is a barrier to scientists’ right to broadly disseminate their findings—Mellman sees copyright as a legal protection for scientists and their work.

#### **Speed slows**

Because scientists and their slow peer review process safeguard the integrity of the information, they must remain heavily involved. “Scientists have a major role to play in ensuring that this movement occurs, but the methods and language that are employed must be tailored to match the complexity of the situation,” says Frank Gannon, Executive Director of the European Molecular Biology Organization.

At the bottom line, “the researcher in search of high-quality information does not need to lose productivity because he or she has to make a lurching adjustment and navigate a new communication system that has been hustled into place prematurely,” says Okerson. That being the case, the speed of the PMC–PLS challenge may slow, not advance science. Ω

*“Scientists need the verification provided by peer review...”*

*“... free on-line access might have the opposite-from-intended effect by pricing some scientists out of the discourse...”*

## Join us on STC's Scientific Communication SIG mailing list!

STC runs an Internet-based e-mail discussion group for the Science SIG. It's a quiet, friendly place to turn for help if you've got questions concerning scientific communication. If you'd like to join, point your Web browser to <http://lists.stc.org/cgi-bin/lyris.pl?enter=stcscsig-L>

There's no cost to join, and you can expect a very low volume of mail. Of course, the more people join, the more traffic there'll be, so please join. It's a great way to make the SIG work for you.

Two other mailing lists of interest to Science SIG members:

**Copyediting-L:** discussions of editing in all its various forms. To subscribe, send the message "subscribe copyediting-L Your name" (with no quotes, and with your actual name instead of "Your name") to [Listserv@listserv.indiana.edu](mailto:Listserv@listserv.indiana.edu)

**Techwr-L:** discussions of the tools and travails of the technical writer. To subscribe, send the message "subscribe techwr-L Your name" (with no quotes, and with your actual name instead of "Your name") to [Lyris@lists.raycomm.com](mailto:Lyris@lists.raycomm.com)

### *the Exchange*

Newsletter of the Scientific Communication SIG

Society for Technical Communication

901 North Stuart St., Suite 904

Arlington, Virginia

22203-1822 U.S.A.