It's more than costs now; it's control of our property. I think universities have realized that what they are producing, whether it's papers or course material, has a value. — *Steve Koonin*, PROVOST AND PROFESSOR OF THEORETICAL PHYSICS

E-Journals:
Do-It-Yourself Publishing

by Jane Dietrich

What's scholarly research worth? A year's subscription (that's 131 issues) to *Brain Research* costs $16,344; *Biochimica Biophysica Acta* (129 issues) will set you back $11,362; the full set of *Physics Letters* (288 issues) goes for $13,843. These journals are all published by Reed Elsevier, the world's largest commercial publisher of science journals.

On the other hand, an institution can subscribe to the weekly *Journal of the American Chemical Society* for a mere $2,165. (And an annual subscription to four issues of *E&S* is still a bargain at $10.)

Why are commercial scholarly journals so expensive? The commercial sector's share of the scholarship market began to expand after World War II as the explosion in scientific information exceeded the capacity of the more traditional publishing avenues—the professional societies. The number of new scientific journals quadrupled between 1940 and 1970, a phenomenon that didn't seem to have a down side until prices began escalating. Between 1986 and 1996, the cost of scholarly journals rose 148 percent (the consumer price index rose 44 percent during the same period), and by the mid-90s librarians had taken notice and were becoming increasingly uneasy. And then prices skyrocketed between 1996 and 1997.

"We were shocked in 1997," says University Librarian Anne Buck, when prices rose 21 percent across the board. Caltech's highest journal increase that year was 29 percent; its lowest about 19 percent. When Buck informed Provost Steve Koonin of the impact on the library's journal budget, a revolt was born. "This is what *we* are producing," thought Koonin, "Why do we have to pay for it?"

As the universities see it, faculty and their research groups, supported by their universities and funding agencies, do the research, write it up, and submit it to a journal's editor or editorial board, which consists of other "volunteer" academic scientists. They send it out to yet another scholar for review (these volunteers are rarely paid). If the paper is accepted, the original scientist makes any required revisions, and eventually it's printed in the journal. Then the universities have the privilege of buying back their own work. Elsevier, responsible for some of the largest price hikes, is making its profit, according to Buck, "on the backs of the libraries and the universities." (Caltech's journal budget stands at around $1.9 million this year, even after the Institute—with faculty collaboration—purged its subscription list of a number of nonessential journals with a high cost per use.)

How do the journal publishers get away with this? Traditionally, an article may be published in only one place, so the publisher has a virtual monopoly on that information. In addition, "the publishers discovered that the market was inelastic," observes Rick Flagan, professor of chemical engineering and former chair of the faculty library committee, "and it's inelastic for a reason: the people who pay the bills and the people who demand the subscriptions are two different sets."

There's also a disconnect between the objectives of the commercial publishers and those of the authors (and the universities that employ them). Scientists want to get their research out to as wide a community as possible, and as quickly and accessibly as possible, whereas publishers are mainly interested in the return on their investment. Professional societies are usually less greedy, and many of them use the profits on their journals to fund member services in their fields. Yet, say critics, those member services should be supported by the societies' members and not by university libraries, who are the ones actually footing the bill.

To rub salt in the wound, the publishers have insisted on holding the copyright, in effect seizing ownership of the intellectual output of the universities. If a scholar at Caltech asks the library to make multiple copies of his own work published

in a journal owned by the library to, say, distribute to his class or send to colleagues, only the first copy is royalty-free. Buck points out that recent proposals to revise American copyright law are pursuing a "one-size-fits-all approach" in an attempt to serve the entertainment world and the software industry as well as publishing. "The entertainment and software businesses want to put up as many barriers as possible to anyone getting their material, particularly in its initial use," she notes. "The problem came when some of the large publishing houses, and basically all of them in scholarly publications, saw this as a great way to piggyback on the process and get a lot of money for themselves—even though their product is not like a movie or a piece of software."

Koonin considers copyright the linchpin of scholarly communication. "The researchers want nothing more than to disseminate their information, yet they are held captive in many ways by the publishers who hold the copyright," says Koonin. "I think that with the electronic media developing, copyright has become a barrier to dissemination of scholarly material rather than the incentive it was originally meant to be." Copyright certainly makes sense for commercial authors, who write for income, but not for researchers, who are only interested in exposure for their work. A licensing agreement allowing the publisher the right to print an article, but not own it, would be fairer, say critics of the current system.

"It's more than costs now; it's control of our property," says Koonin. "I think universities have realized that what they are producing, whether it's papers or course material, has a value. Universities are 'content providers,' and the new electronic media opened ways of disseminating that content in much more cost-effective ways than we could before." Print journals have not ignored the potential of the Internet, and many of them do have on-line versions; these can usually be had, though, only by "pay-per-view."

Librarians and scientists both fear that if a commercial journal exists *only* in electronic form, the unprofitable archives of back files would not be a high priority and might get dumped. Not only that, but when your on-line subscription expires, so does your access. In contrast, if you don't renew a print journal, the back issues you've already paid for are still yours to keep.

But e-journals have a lot of advantages over printed publications: papers can be disseminated almost instantaneously; all sorts of search options are available; papers on similar topics can be retrieved through links; references and an author's previous work can be linked to the current paper; comments and comments on comments can be hooked onto a paper; video, sound, 3-D graphics, and data sets can be incorporated into the text. Even something as simple as color images can be added without the extra charges that publishers currently impose. So it isn't only the economics of journal pricing that is prompting the revolution; the technology is already there and waiting for it.

Caltech's first action was to convene a conference on "The Future of Scholarly Communication"

On his paperless (at least for the photo) desk, Provost Steve Koonin logs onto the no-frills physics e-print archive of xxx.lanl.gov. Koonin posts all his research papers on the site before publication in a paper journal.

A 21-percent increase in the price of scholarly journals in 1997 helped motivate University Librarian Anne Buck to look at how to tap technology. She is coauthor of the proposal for *Scholars Forum*, a model for a multidisciplinary electronic journal.

When you go into an entirely electronic world, there are some very serious issues in my mind about what happens to the past, what happens to the record of scholarship. — *Anne Buck*, UNIVERSITY LIBRARIAN

in March 1997. Attended by 55 representatives of 29 universities, the conference featured four speakers who were prominent proponents of electronic publishing, as well as two panels—one of university provosts and one made up of representatives of professional societies.

"We brought together people who had the power to make decisions," says Flagan, an avid advocate of the e-revolution, "some librarians, but also people who oversee the library budget and who are motivated to see something happen. The questions that were addressed at that conference were basically: What is it that universities need to communicate for the future? What would we do with a clean slate? Suppose there were no journals today, and suddenly this thing called the Internet came along and we wanted to do something to communicate the results of our research? How would we do it if we were starting from scratch?"

Journals do, of course, contribute some value for the money: they provide mechanisms for editing, for distribution, for easy access to information, for preservation of the scholarly record, and for certification. Certification, in the form of peer review, is critical to the functioning of research universities, and it's what gives the journals their enormous clout. This stamp of approval on someone's work determines who gets hired, who gets tenured, who gets promoted. But the main insight to come out of Caltech's 1997 conference is that peer review is not inherently tied to a print journal. Academics are doing this work for free anyway; they could just as well do it in another kind of distribution system, say an electronic one, if universities agreed to stand together to accept this

stamp of approval. Koonin is credited with advancing this notion, henceforth referred to as "decoupling" the refereeing from the journals. And it goes further: you can also decouple the editorial function and the archiving.

The librarians (who have to pay the journal bills) and the provosts (who have to come up with the money for the librarians) left the conference inspired by the prospect of decoupling.   The group most dependent on the journal system, however—the faculty—was not so easy to convince. If tenure and promotion are tied up with the old system, who would want to take a risk on something new and unknown? Many believe that what the journals provide—vetted and edited papers aggregated into neat little packages as the traditional ticket to tenure—is worth the cost to libraries, as well as the price of giving up copyright.

But in a faculty meeting in the spring of 1998, Koonin issued a challenge. He noted that Caltech already requires faculty members to sign an agreement that all patents and copyrights that result from their Institute- or grant-supported work belong to Caltech. No one had ever thought much about copyright, and that part of the agreement was never enforced. But suppose we started enforcing it? mused Koonin and then suggested that Caltech authors withhold copyright from the publishers. No groundswell emerged of professors eager to attempt this, and Caltech, of course, never followed through on the implied threat. Koonin, at least, has practiced what he preached, and one journal to which he contributed subsequently changed its copyright policy in response to his raising the issue.

The copyright challenge did rouse a few adherents at both poles—those enthusiastic about taking on the journals and those who fervently believed that the present system worked just fine—with the vast majority of the faculty indifferent to the entire issue. Professor of
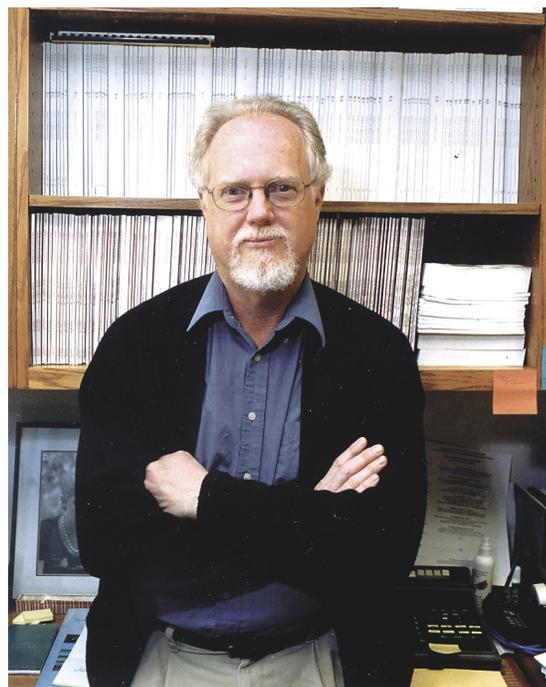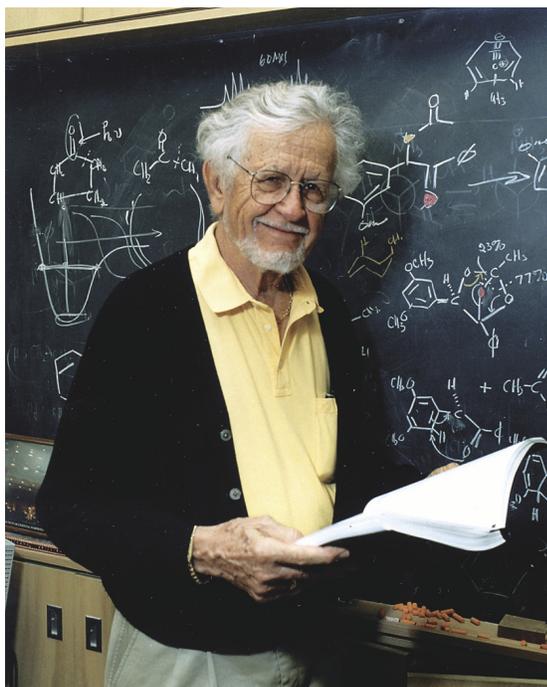
Planetary Science Bruce Murray set up an on-line, "threaded" discussion system—a "hyperforum"—to discuss copyright and the question: "Will the accelerating trends toward electronic publishing and Internet commerce overturn traditional relationships between university researchers, publishers, and the scientific endeavor?" Over the three months that it was up in the fall of 1998, the hyperforum attracted log-ons from only 40 members of the faculty, 16 grad students, 1 postdoc, and 40 members of the staff. Of these, only 16 posted comments.

While you might think that those who had the most to lose by shaking up the system on which their tenure depends—the junior faculty—would be the most passionate about leaving things as they are, it was Institute Professor of Chemistry, Emeritus, Jack Roberts who led the defense. He contributed numerous comments to the hyperforum, arguing that everything the journals provided was well worth giving up his personal ownership of copyright, and that it was unreasonable and naive to expect the journals to do all the work of publishing and then allow anyone to copy articles or disseminate them on the Web for free.

Roberts also touched on the importance of permanent archives and the differences among disciplines. "There seems to be an operational feeling in physics that anything that is older than a few years is of little interest, except as history," he wrote. "Chemistry is different. Chemists need all kinds of tidbits of information, particularly about preparations, that can be supplied by papers published more than a century ago."

Facetiousness also crept into the hyperforum. Under the title "A Brave New World," Roberts posted a mock news release announcing the new "all-electronic, World Wide Web-based *California Institute Journal for Engineering, Science, Humanities and Social Sciences* (*CALJESHSS*), edited by the Institute's own B. C. Murray, S. E. Koonin, and R. C. Flagan" which would be "free of all of the restrictions which for two centuries have cramped the style of the authors of old-fashioned research journals as to time taken for reviews, length, copyrights, number of illustrations . . . audio records and animations." He went on to describe the wonders of the electronic future and ended with the news that "*CALJESHSS* is developing hardware and software so that research can be published that involves direct transmission of research data on odors, tastes and tactile responses."

> There seems to be an operational feeling in physics that anything that is older than a few years is of little interest, except as history. Chemistry is different. Chemists need all kinds of tidbits of information, particularly about preparations, that can be supplied by papers published more than a century ago.
>
> — *Jack Roberts*, INSTITUTE PROFESSOR OF CHEMISTRY, EMERITUS

David Goodstein, the Gilloon Professor and professor of physics and applied physics, as well as vice provost, responded that *CALJESHSS* sounded like "a magazine devoted to the Second Coming in California" and suggested *The Beaver Dam* instead.

Neither name, fortunately, was destined to catch on. In the meantime, Buck and Flagan (along with Betsy Coles, manager of digital library systems) had come up with their own title and full-fledged proposal, which they posted last March on the library's Web site: "Scholars Forum: A New Model for Scholarly Communication." The site <http://library.caltech.edu/publications/ScholarsForum> has been getting about 1,500 hits a month. *Scholars Forum* lays out a plan for a dual system consisting of a multidisciplinary database of papers, including preprints (posted by the authors) and certified papers (posted by editorial boards) that have successfully undergone peer review. These final papers could be compiled into electronic journals and would remain accessible through electronic indexing and archiving.

Buck and Flagan don't advocate that Caltech go it alone, but envision a trilateral partnership between a consortium of universities, the professional societies, and the authors themselves. Professional societies within the various disciplines would continue to maintain editorial boards to validate papers and distribute them in print or electronic form, but other editorial boards could also spring up under the aegis of the *Scholars Forum* consortium. None of these boards would be granted exclusivity, and authors would retain copyright.

The proposal considers universities and univer-sity libraries the natural choice to control and archive the work that they produce, and endows the university consortium with the responsibility for maintaining the servers, developing and maintaining operating standards and protocols, and supporting the preservation of the scholarly record. "The operator of the server should be an institution that has a likelihood of long life," states Flagan. "There has to be more than one server, and there has to be a commitment by the operators of the servers to translate as technology changes. It takes people, it takes computers, it takes institutional memory. So you want the people who have the commitment to do it, and they are the institutions that support the sciences."

Who would pay for this? "Who benefits the most from publishing?" asks Flagan. "The author and the author's institution." So they (preferably the institution, he adds) should bear the costs, which wouldn't be large, although no one really knows yet what something like the *Scholars Forum* will cost. But many journals already require page charges from authors, so it wouldn't be unreason-able to ask the equivalent of page charges to sup-port the office and secretarial costs, as well as the cost of putting refereed papers on the server. And the *Scholars Forum* suggests that the author pay for copy editing and for any necessary writing assistance.

Is Caltech willing to back up something like the *Scholars Forum* with funding? Says Koonin: "A part of what we're supposed to do in a univer-sity is promote the dissemination of knowledge, and I would much rather pay whatever it costs for us, the universities, to put it on the Net for free, worldwide access, than pay some commercial publisher or even a society for one or two copies."

As for the business of indexing and archiving, "the logical thing would be to say this is a new role for libraries," Flagan adds. "The libraries have traditionally been the holders of the print archive; let's make them the holders of the electronic archive."

Librarians, haunted by the incineration of much

> The publishers discovered that the market was inelastic, and it's inelastic for a reason: the people who pay the bills and the people who demand the subscrip-tions are two different sets.
>
> — *Rick Flagan*, PROFESSOR OF CHEMICAL ENGINEERING

of the written knowledge of the ancient world when the great library of Alexandria burned a couple of thousand years ago, take this very seri-ously. "When you go into an entirely elec-tronic world," remarks Buck, "there are some very serious issues in my mind about what happens to

the past, what happens to the record of scholarship. This is especially true in science, because science is a cumulative process." To prevent another catastrophe of Alexandrian proportions (until the long-term retention of electronic files is secure), the *Scholars Forum* proposes that a few copies of the "journals" still be deposited in widely dispersed libraries on acid-free paper guaranteed to last 200 years.

"The nice thing about paper," notes Eric Van de Velde, director of Caltech's Library Information Technology Group, "is that benign neglect works. You put it in a room and you forget about it. It gathers dust, but it will remain readable." But how *will* you store a digital library? For example, CD-ROMs disintegrate in 15 to 20 years, says Van de Velde, but they'll be obsolete soon anyway, taken over by DVDs. And although DVD players can still read CD-ROMs, will the next generation of technology be able to read them? The same is true of other technologies. Benign neglect isn't going to work here, according to Van de Velde. "The digital archive must be actively managed." An electronic format also makes it possible to publish the raw data of experimental results, video, three-dimensional structures; how will that be stored? "There are so many different things that you could store," says Van de Velde, "but how do you insure that Microsoft Word 700, or 3000, or whatever, can still read it?"

I can easily imagine a future where every university is basically the stakeholder

of its intellectual information, keeps track of the papers it produces, and gives

access to people under policies that it decides. — *Eric Van de Velde*, DIRECTOR,

LIBRARY INFORMATION TECHNOLOGY GROUP

Van de Velde, who moved over to the library from applied mathematics, does believe that this will be possible, "but difficult." Buck, as an information manager, worries about "chaos in the record in the interim." What happens to articles that are published during the period of transition, when there is no way to preserve them to guarantee that they can be read in the future? That period could last a generation, she believes.

But before you can even think about storing them, e-prints must be collected into a uniform database, and the other principal technical problem is developing common protocols or formats for submission. "Each discipline has its way of producing its manuscripts," observes Van de Velde. "You have many different possibilities for submitting manuscripts, and somehow the same system has to be able to handle them with as little human intervention as possible. Right now, we need to be able to support a wide variety of formats," he adds.

Neither the Conference on Scholarly Communication nor the *Scholars Forum* dealt in depth with the technological nuts and bolts. Fortunately, there already is a flourishing prototype of such a system. In 1991 Paul Ginsparg at Los Alamos National Laboratory created a self-archiving preprint server <xxx.lanl.gov> for high-energy physics, where scientists can post their papers, or "e-prints." It has now expanded to all of physics, as well as astronomy and mathematics, and holds more than 100,000 records in its database (and claims over 50,000 users daily). Most physicists assert they couldn't live without it, and journals have come around to coexisting with the site, since without peer review, they can rationalize that this isn't really "publication," so the article can still be published in a journal. Koonin says that "for the last decade all of my papers have been posted there at the same time that they're submitted to a journal." Ginsparg has proved e-publishing can be done and be very successful.

The National Institutes of Health has proposed something similar for the biomedical sciences. Originally called *E-biomed*, it would have a governing board of scientists and consist of two sections—one for peer-reviewed papers (which would be done by the relevant scientific societies) and another for unreviewed e-prints. Former NIH Director Harold Varmus touted the proposal for providing instantaneous, cost-free access to research, which would accelerate the exchange of information among scientists. But the proposal caught a lot of flak, mainly from medical journals published by medical societies and commercial publishers, who were decidedly unenthusiastic. Critics point to the dangers of concentrating too much power in a governmental agency and of allowing public access to unvetted medical information. Undaunted, the NIH planned to put its electronic archive, renamed *PubMed Central* and expanded to encompass all the life sciences, on line in January. "Biomedical is the big gorilla here,"

says Koonin, who believes things may change rapidly if this venture succeeds.

Even as electronic publishing is catching on within particular scientific fields, visionaries want to extend these models to all disciplines. The Open Archives Initiative (formerly known as the Universal Preprint Service Initiative) aimed to do just that in its first meeting in Santa Fe this past October: link the archives of many disciplines "to ensure that they work together so that any paper in any of these archives could be found from anyone's desktop worldwide, as if it were all in one virtual public library." Representatives, mostly digital-library experts and computer scientists, from universities, libraries, and various electronic publishing undertakings attempted to combine their knowledge into a usable system. Van de Velde represented Caltech.

"Interoperability" was the key word: how to structure the various kinds of archives that are likely to emerge from different institutions and different fields so that they are universally accessible to the end-user. Some initial standard mechanisms and technical requirements were formalized by the Open Archives Initiative as the Santa Fe Conventions, which will be implemented in already existing archives over the next six months.

Van de Velde served on a panel discussing the pros and cons of institutional archives versus those oriented around disciplines. He advocates the former. "I can easily imagine a future where every university is basically the stakeholder of its intellectual information, keeps track of the papers it produces, and gives access to people under policies that it decides. If each university maintains such a database, the universities can link them in a way that you can search them all. Organizationally,

this would be very clean and straightforward. Realistically, however, we must expect and plan for archives by many other organizations, such as publishers and societies."

Van de Velde is confident that with enough people working on it, this can all be done. "Everybody knows it can be done, and in many respects it already *has* been done. But you can always make it easier to use, and that is definitely important for widespread user acceptance."

Acceptance remains the biggest hurdle. Who would want to publish in a *Scholars Forum* or in some vast anonymous archive if they could publish their paper in, say, *Brain Research* or *Physics Letters* instead? Many peer-reviewed, start-up electronic journals have experienced credibility problems. How do you establish a reputation? Will electronically published papers count for tenure? And judging from the lack of interest in the hyperforum, it will be tough to persuade many Caltech faculty members that this is the way to go.

The nature of publishing seems to be changing inexorably, like it or not, and Van de Velde believes that electronic publishing has "a very, very high probability of success, because scientists do want access to the literature in easy electronic form. We can actually see it in the library here. For the last few months we have been able to provide some documents through Ibid, a Caltech electronic document-delivery service. A big majority of users prefer electronic document delivery, and even though we can't do everything electronically yet, you can see that the electronic format is important to researchers here."

But there's still a lot of convincing to do, and the revolution will progress in increments. Van de Velde and his Library Information Technology Group (which consists of six people, none of whom are librarians) are currently installing various kinds of software including NCSTRL (National Computer Science Technical Report Library) and NDLTD (Networked Digital Library of Theses and Dissertations), a system developed at Virginia Tech for submitting dissertations electronically. The *Caltech Undergraduate Research Journal* (CURJ) is looking into going on line. And another likely candidate for early digitizing is conference proceedings, say the creators of the *Scholars Forum* proposal.

Whether Caltech's do-it-yourself entry into electronic publishing will actually take the form proposed in the *Scholars Forum* is debatable. "It's a framework," says Buck, "but I don't think it's going to be taken really seriously by the science community until we have some findings. It's suffering from a lack of 'and here's what it looks like.'"

"It's an interesting model," says Koonin. "It's caught some attention among the people who are interested in this. My guess is that the final system won't look exactly like that, but it's a good start." □