Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications: Proceedings of a Symposium
Committee on Electronic Scientific, Technical, and Medical Journal Publishing, National Research Council

ISBN: 0-309-53230-2, 136 pages, 8 1/2 x 11, (2004)

This free PDF was downloaded from:
http://www.nap.edu/catalog/10983.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Purchase printed books and PDF files
- Explore our innovative research tools – try the Research Dashboard now
- Sign up to be notified when new books are published

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

ELECTRONIC SCIENTIFIC, TECHNICAL, AND MEDICAL JOURNAL PUBLISHING AND ITS IMPLICATIONS

PROCEEDINGS OF A SYMPOSIUM

Committee on Electronic Scientific, Technical, and Medical Journal Publishing

Committee on Science, Engineering, and Public Policy
Policy and Global Affairs Division

THE NATIONAL ACADEMIES

THE NATIONAL ACADEMIES PRESS
Washington, D.C.
www.nap.edu

THE NATIONAL ACADEMIES PRESS   500 Fifth Street, N.W.   Washington, DC 20001

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chair and vice chair, respectively, of the National Research Council.

www.national-academies.org

iii

STEERING COMMITTEE ON ELECTRONIC SCIENTIFIC, TECHNICAL, AND MEDICAL JOURNAL PUBLISHING AND ITS IMPLICATIONS

EDWARD H. SHORTLIFFE (Chair), Columbia University Medical Center
DANIEL ATKINS, University of Michigan
FLOYD BLOOM, The Scripps Research Institute
JANE GINSBURG, Columbia University School of Law
CLIFFORD LYNCH, The Coalition for Networked Information
JEFFREY MACKIE-MASON, University of Michigan
ANN OKERSON, Yale University
MARY WALTHAM, Publishing Consultant

Principal Project Staff

Paul Uhlir, Project Director
Alan Inouye, Senior Program Officer
Julie Esanu, Program Officer
Robin Schoen, Program Officer
Kevin Rowan, Project Associate
Amy Franklin, Senior Program Assistant

COMMITTEE ON SCIENCE, ENGINEERING, AND PUBLIC POLICY

MAXINE F. SINGER (Chair), Carnegie Institution of Washington
R. JAMES COOK, Washington State University
HAILE T. DEBAS, University of California, San Francisco
MARYE ANNE FOX, North Carolina State University
ELSA M. GARMIRE, Dartmouth College
MARY-CLAIRE KING, University of Washington
W. CARL LINEBERGER, University of Colorado
ANNE C. PETERSEN, W. K. Kellogg Foundation
CECIL B. PICKETT, Schering-Plough Research Institute
GERALD M. RUBIN, Howard Hughes Medical Institute
EDWARD H. SHORTLIFFE, Columbia University Medical Center
HUGO F. SONNENSCHEIN, The University of Chicago
IRVING L. WEISSMAN, Stanford University, School of Medicine
SHEILA WIDNALL, Massachusetts Institute of Technology
MARY LOU ZOBACK, U.S. Geological Survey

Staff

Richard E. Bissell, Executive Director
Deborah D. Stine, Associate Director
Marion E. Ramsey, Administrative Associate

v

PREFACE

The use of the Internet and other digital information technologies by the scientific, technical, and medical (STM) research community in the United States and most other countries has transformed many aspects of the research and publishing process. The new technologies have created fundamental changes in the production, management, dissemination, and use of all types of information. It is now possible to communicate research results much more quickly, broadly, and openly than was possible through traditional print publications in the past. Researchers are now able to make available independently their data and articles online, where the information may be easily found, browsed, annotated, critiqued, downloaded, and freely shared. This is resulting in significant changes to the linear path of writing, refereeing, and reviewing of publications as all these functions can be performed concurrently. Most STM publishers also now publish electronic versions of their journals, some exclusively so. The technological developments and resulting changes to the sociology of science are creating both opportunities and challenges for the effective management of scientific communication generally, and STM publishing more specifically.

Because of the far-reaching implications of these developments, the National Academy of Sciences Council's Committee on Publications recommended that the council commission a study of the factors involved in the changing mechanisms for access to STM information in the scholarly publications and the various technical, legal, policy, and economic issues that they raise. The committee indicated that it is imperative for the National Academies to address, in particular, the increasing concerns about the implications of various models for access to STM publications for the scientific community.

As a result, the Committee on Science, Engineering, and Public Policy was asked to appoint a committee to oversee the planning for the *Symposium on Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications,* which was held May 19-20, 2003, at the National Academy of Sciences in Washington, D.C. The symposium brought together experts in STM publishing, both producers and users of these publications, to: (1) identify the recent technical changes in publishing, and other factors, that influence the decisions of journal publishers to produce journals electronically; (2) identify the needs of the scientific, engineering, and medical community as users of journals, whether electronic or printed; (3) discuss the responses of not-for-profit and commercial STM publishers and of other stakeholders in the STM community to the opportunities and challenges posed by the shift to electronic publishing; and (4) examine the spectrum of proposals that has been put forth to respond to the needs of users as the publishing industry shifts to electronic information production and dissemination.

The symposium was divided into six sessions, each introduced by opening comments from a moderator, followed by several invited presentations. Session 1 examined the costs involved with the publication of STM journals while Session 2 looked at the related publication business models. Session 3 explored the legal issues in the production and dissemination of these journals. Sessions 4 and 5 looked toward the future and examined, respectively, what is publication in the future and what constitutes a publication in the digital environment. The final session provided several commentaries on the results of the symposium.

This publication presents the *Proceedings* of the symposium. The speakers' remarks were taped and transcribed, and subsequently edited. The statements made in the enclosed papers are those of the individual authors and do not necessarily represent the positions of the steering committee or the National Academies. The National Academies hosted a live audio Webcast of the symposium to reach a broad audience and receive additional input. This Webcast can be found on the symposium Web site at: http://www7.nationalacademies.org/cosepup/E-Publishing.html/. A summary report prepared by the symposium committee has been published separately and is available from the National Academies Press.

Edward Shortliffe   *Paul Uhlir*
Committee Chair   *Project Director*

ACKNOWLEDGMENTS

Contents

Appendixes

1. INTRODUCTORY REMARKS

Bruce Alberts, National Academy of Sciences

This symposium is part of a longstanding effort of the National Academies to promote wise policies for science. The important functions of scientific publication from a policy perspective include scientific validation, the dissemination of information, and the specific obligations created when an author publishes.

The scientific validation function of publication is accomplished by peer review and by editing to ensure that the science is sound. These processes also ensure that the data and the methods are complete and clearly presented. This is what distinguishes the information in scientific journals from the vast amount of other material that is also on the Internet. There are many publications that make a false claim to be scientific, and since the seventeenth century the scientific community itself has organized to discriminate between what is, and what is not, good science. Without this discrimination, science could not move steadily forward by building upon what has already been discovered.

Publication also has a critical dissemination function. That is what is largely motivating this symposium, because the Internet is a radically new way to disseminate scientific information. This powerful communication channel greatly enhances the potential reach of science; it allows us to include nations and people who would otherwise not be reachable. Equally important, by creating powerful search engines that can rapidly find desired information, we can make much better use of the vast store of data and scientific information available.

We could do a lot better, however. For instance, there are great opportunities for data mining that have not yet been adequately exploited. In addition, there are many journals that are still not available in electronic form. Moreover, foreign journals are not comprehensively abstracted by U.S. services such as PubMed. Science is an international activity, and we need to make the knowledge that is developed everywhere more readily available to other scientists.

Our own journal is the *Proceedings of the National Academy of Sciences* (PNAS). We have tried to set an example by making its electronic version as widely available as possible. The entire 16,000 pages each year of PNAS are immediately available free on the Internet to those in most developing nations, and this information can be freely accessed by everyone after a six-month delay.

The National Academies publish all of our reports through the National Academies Press. This is a different kind of scientific literature—the results of consensus studies on issues such as the health effects of arsenic in drinking water, the science of climate change, and thousands of other topics. There are about 10 years' worth of National Academies reports—some 2,900 books—online, all readable page by page. In addition, the PDF files for these reports are made freely available to anyone in 130 developing nations.

What about the obligations of authors? The National Academies held a workshop in February 2002 that was sponsored by the Board of Life Sciences entitled "Sharing Publication-related Data and Materials: Responsibilities of Authorship in the Life Sciences." Tom Cech, the head of the Howard Hughes Medical Institute and a Nobel Prize-winning scientist, was the chair of the committee that organized this event. The final report from the workshop contained the following statement:

> The publication of scientific information is intended to move science forward. More specifically, the act of publishing is a *quid pro quo* in which authors receive credit and acknowledgement in exchange for disclosure of their scientific findings, providing them a forum on which other scientists can build with further research. An author, therefore, has the obligation to release both data and materials to enable others to verify and extend published findings.

2

Publication therefore creates important obligations on the part of the authors that need to be enforced, both through the grant agencies that fund the research and through the journals that publish the work. It is only with this kind of sharing, enforced by the scientific community, that the public's generous support of knowledge production through scientific research can produce its intended benefits.

2. KEYNOTE ADDRESS

James Duderstadt, University of Michigan

The goal of this symposium is to bring together experts from an array of constituencies, including both producers and users of STM publications, to look at some of the technical changes that have occurred in electronic publishing, and how they influence decisions to publish or not; to identify the needs of the science, technical, and medical (STM) publishing enterprise itself as users of journals; to understand the responses of both the commercial and not-for-profit STM publishers; and to examine a very broad spectrum of proposals and activities underway that are attempting to respond to the needs of the community with these new technologies.

A major focus of this symposium is to look at business models and to try to establish the degree to which they address many of the challenges and concerns. During the discussions it is important to keep in mind the ongoing developments in the scientific enterprise itself, stimulating in part, and being stimulated by, this kind of scholarly communication.

How is electronic publishing affecting the practice of scientific research—the communication of research results to scholars and others, perhaps including the public; the curation of data and evaluation of research; and archiving of results? The challenge is to identify the issues and problems that the STM community needs to control and resolve if it is to exploit the remarkable opportunities presented by this very rapidly evolving technology and also cope with the challenges it presents.

The current situation can perhaps be described as a chaos of concerns, with the continuation of some disturbing trends that have evolved over the last couple of decades. Access to STM information is increasingly expensive, and in some cases restricted. And yet, the amount of information generated at research institutes continues to grow.

Journal subscription prices continue to escalate, yet university library budgets fail to keep pace, particularly in these days of economic challenges at both the federal and state level. The price inflation in electronic publication resources, estimated to be running at approximately 10 percent per year over the past decade,[1] has continued to run well ahead of the consumer price index. But even more dramatic has been the increase in the pricing for reference tools, increasing as much as 600 percent over the print cost of bound volumes. To these challenges should be added the growing complexity of dealing with various financial models and the licensing schemes that provide access.

It is clear that these new technologies have created very fundamental changes in the production, management, dissemination, and use of all kinds of information. If one were to categorize very simply the two camps of concerns, on the part of the publishers the critical question is, How many copies of work will be sold or licensed if networks make possible planetwide access? And the nightmare, of course, is that the answer is only one. One document can be replicated time and time again, to not only serve, but perhaps collapse, the entire marketplace.

On the other side, the nightmare to consumers is that in our efforts to preserve the marketplace, we will put in place an array of technical and legal protections that reduce access to what should be a public good, society's intellectual and cultural heritage.

There are a lot of reactions and counter-reactions at the university level. The first reaction is a budgetary one. University libraries simply cancel many subscriptions. In some cases this is mandated by limited resources. In other cases it is an effort to get the attention of the publishing industry, although what cancellation generally does is simply drive up the costs even further for the remaining subscriber base.

---

[1] See Barbara Albee and Brenda Dingley. 2003. "U.S. Periodical Prices—2002," American Library Association, Washington, DC. Available at
http://www.ala.org/Content/NavigationMenu/Products_and_Publications/Periodicals/American_Libraries/Selected_articles/U_S__Periodical_Prices,_2002.htm.

In other instances, we have seen rebellion at the grassroots. Editorial boards have protested against the commercial publisher journal prices and have resigned and moved to less expensive publishers in scientific societies. The complexity and shifting from a first-sale approach characteristic of paper to licensing have caused a good deal of experimentation.

There are many other variants. One approach is reminiscent the way that dissertations used to be generated from the university microfilms collection, as an edition of one. That is, to begin to make it acceptable that there may only be one physical copy of a document but have the ability to reproduce that from an online copy, at the user's expense.

Another interesting approach that is emerging involves the open source or open-access strategy. The success of the open source software movement through Linux, the Apache Web server, and similar technology has given rise to a number of related open-access initiatives, such as the Open Knowledge Initiative and the MIT OpenCourseWare project. These initiatives focus on developing new financial models for the open distribution of scholarly materials, perhaps by building charges for dissemination into research grants that generate the information in the first place. This is not only consistent with the traditions and values of academia but also reinforces the definition of the university as a public good, an issue that university leaders are increasingly worrying about these days, when the rest of society tends to look at us more as a market commodity.

In summary, advances in digital technology are producing radical shifts in our ability to reproduce, distribute, control, and publish information. Yet, as these advances become more a part of scientific activity, they tend to run headlong into the existing practices, policies, and laws that govern traditional publishing.

The issues are complex, in part, because the stakeholders are so many, so varied, and with different agendas. People who fund research want to see that the information is advanced and made available to the public. The authors, editors, and reviewers do not charge for their labor. They are motivated to contribute to the public good, but of course they also have other rewards, not the least of which is tenure. Publishers, as intermediaries, although they do not pay for content, do add significant value and provide the work in published form. Libraries, similarly, are intermediaries. They provide access to the users of STM content. They pay the subscription fees, but they usually do not charge for providing access. And, of course, the end users either pay for personal subscriptions or obtain the resources free through libraries.

There are several more general issues that need to be considered. First, is the changing nature of science and technology research. As pointed out in the recent National Science Foundation (NSF) report, *Revolutionizing Science and Engineering Through CyberInfrastructure*,[2] the process of knowledge creation itself—experimentation, analysis, theory development, and forming conclusions—is increasingly occurring entirely in the digital world. That has caused a shift from the sequential process of research, publication, validation, and dissemination to more of a parallel flow model that is interactive, in which the process of publication and distribution actually becomes almost the process of research itself. The key point of the report is that distributed network computing technology is providing a new kind of infrastructure for federating people, information, computational tools and services, and specialized facilities into virtual organizations—so-called collaboratories or grid communities or, as the Europeans call it, e-science, a cyberinfrastructure. The vision put forth by this NSF report is to use this infrastructure to build ubiquitous and comprehensive digital environments. Such environments will become interactive and functionally complete for research communities in terms of the people, data, information tools, and instruments, and that operate at unprecedented levels of computational storage and data transfer capacity. Part of the aim is to trigger the necessary public and private investments to create this cyberinfrastructure. Nevertheless, many elements of it are already in place, and it will significantly change the nature of scholarly activity, including scholarly publication.

The reality today is that electronic publishing is becoming the dominant mechanism for publishing and reading scholarly materials. It opens vast possibilities, of course, but it challenges existing practices and principles, including the way in which we handle intellectual property. A new paradigm for scholarly

---

[2] National Science Foundation. 2003. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, Arlington, VA, January.

communications is coming into focus that is capable of providing open online access to the work of scholars without payment; online repositories of high-quality, certified materials; and a stable economic model to sustain these resources.

This will pose a particular challenge to libraries, shifting them from a focus on collecting and archiving knowledge resources, to assisting scholars in navigating these resources. Today, the campus library has become somewhat less central to researchers' lives. The library has evolved from a place into a utility. It too is becoming a part of the Internet.

Legal and policy issues are the second major issue area. It is clear that the emerging digital infrastructure imperils a great many of our existing practices, policies, and laws that have served intellectual activity in this country and globally so well over the past two centuries, forcing the rethinking of some fundamental premises and practices associated with intellectual property. Indeed, there is a concern that many of these will be challenged to the bedrock.

The third topic concerns the evolution of digital technology. In 2000 the National Academies created a study group chosen from industry, higher education, and federal policy development to understand better what the implications of digital technology were for the research university, and even more broadly, for the research enterprise.[3] The concern was that although the opportunities and challenges of this technology were important, many of the most significant issues were neither well recognized nor understood. Among the early conclusions of this effort was that the recognition that the extraordinary evolutionary pace of digital technology shows no sign of slowing, with some aspects such as storage and wireless bandwidth evolving at superexponential rates

The second conclusion was that the impact of the technology on the university will be profound, rapid, unpredictable, discontinuous, and disruptive. It will affect all of the activities of the university– teaching, research, outreach, its organization, financing, governance, even the definition of its faculty and students. Procrastination and inaction are the most dangerous courses of all during a time of rampant technological change.

The report's third major conclusion, and an interesting one, was that universities should begin the development of strategies for facing this kind of technology-driven change with a firm understanding of those key values, missions, and roles that need to be protected and preserved during a time of transformation. These include traditions such as openness, academic freedom, and the rigorous of academic inquiry.

A fourth area of concern is the commercialization of academic output, as the soaring commercial value of much of the intellectual property produced on the campuses raises very significant challenges to traditions such as openness and academic freedom.

Finally, there are the issues of national security, which again call into the question of balancing scientific openness and with the restrictions on public information necessary for homeland security.

As we address these complex issues, we might well keep in mind the well known observation of Thomas Jefferson:

> If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess, as long as he keeps it to himself. But the moment it is divulged, it forces itself into the possession of everyone, and the receiver cannot dispossess himself of it.
>
> That ideas should freely spread from one to another over the globe for the moral and mutual instruction of man, and the improvement of his condition, seems to have been peculiarly and benevolently designed by nature when she made them like fire, expansible over all space without lessening their density at any point, and like the air in which we breathe, move and have our physical being, incapable of confinement or exclusive appropriation. Inventions then cannot, in nature, be a subject of property.

---

[3] See National Research Council. 2003. *Issues for Science and Engineering Researchers in the Digital Age*, National Academy Press, Washington, D.C.

5

3. COSTS OF PUBLICATION

INTRODUCTORY COMMENTS

Floyd Bloom, The Scripps Research Institute

Cost concerns control quality, timeliness, and access so it is natural to begin the symposium with the consideration of the costs of scientific and medical publishing. We have to consider the costs of production, the costs of paper and ink, the costs of dissemination, the costs of acquiring the content that will be produced, and those costs that make a journal competitive with other journals. We also have to be concerned with the explicit, rigorous peer-review processes of scientific journals and the quality of what is published.

As we enter into the era of online publication, we encounter additional costs and concerns. How about the linking to the databases that give a publication equal access and immediacy, and carry it back into the perspective of the past? Why do these things cost so much, and which of these costs could we factor out and control if we knew how to?

The panel that addressed these issues included a variety of perspectives: a librarian, publishers of online and print journals of large and medium-sized societies, and a commercial publisher.

OVERVIEW OF THE COSTS OF PUBLICATION

Michael Keller, Stanford University

Michael Keller, Stanford University librarian and chairman of the board of HighWire Press, began the session by presenting some recent statistics compiled by Michael Clarke of the American Academy of Pediatrics, which provide a good overview of the marketplace for scientific, technical, and medical (STM) journals. Although these statistics are not directly on the cost of publications of electronic journals, they do set the stage for many of the issues discussed below.

According to a recent Morgan Stanley Industry Report,[4] the STM journal market has been the fastest growing segment of the media industry for the past 15 years, with 10 percent annual growth over the past 18 years, and a predicted 5 to 6 percent annual growth over the next 5 years. At the same time, there has been a consolidation of STM journals into a small number of giant publishers. More than 50 percent of STM journals are published by the 20 largest publishers.

According to the Ingenta Institute,[5] until recently, the number of scientific journals doubled every 15 years but the rate of growth of new titles has substantially slowed over the past 5 years; journal size has grown in order to compensate for the lack of new titles; cost of subscriptions has escalated due to size of issues and investments in online technologies; and library budgets are under increasing pressure. There has been a proliferation of library consortia over the past 25 years. These consortia are playing an increased role in electronic journal purchasing decisions and negotiations. The "big deal" began in 1997. It guarantees publishers steady income and provides libraries with steady prices and increased access to titles. The majority of big deals have been signed in the past three years, and the first large wave of contracts will be expiring in 2003. Predictions for the institutional marketplace include decline of library materials budgets, decline of the big deal, a move toward smaller collections, selection of titles will return to importance, and there will be animosity toward for-profit publishers.

---

[4] Morgan Stanley Industry Report. 2002. *Scientific Publishing: Knowledge Is Power*, New York, NY.
[5] Ingenta Institute. 2002. *The Consortium Site License: Is It a Sustainable Model?* Oxford, UK.

Main Elements of Expense Budgets for Some STM Journal Publications

First, the processing costs for the content (i.e., articles, reports of the results, and methods of scholarly investigation) of STM journals come in several subcategories:[6] (1) manuscript submission, tracking, and refereeing operations; (2) editing and proofing the contents; (3) composition of pages; and (4) processing special graphics and color images. Internet publishing and its capacity to deliver more images, more color, and more moving or operating graphics have made this expense grow for STM publishers in the past decade.

The second category of expense is a familiar one, but is also one of two targets for complete removal from the publishers' costs—the costs of paper, printing, and binding, as well as mailings. As researchers educated and beginning their careers in the 1990s replace retiring older members of the STM research community, publishers might finally switch over to entirely Internet-based editions and distribute no paper at all. This transformation would thus move the costs of printing and paper to the consumer desiring articles in that form and remove binding and mailing from the equation altogether.

A third category of expense is that of the Internet publishing services. These are new costs, and they include many activities performed mainly by machines, though in some situations staff perform quality control pre- and post-publication to check and fix errors introduced through the publishing chain. The elements of these costs vary tremendously among publishers and Internet publishing services. At the high end they can include parsing supplied text into a rigorously controlled version of SGML or XML; making hyperlinks to data and metadata algorithmically; presenting multiple resolutions of images; offering numerous elaborate search and retrieval possibilities; supporting reader feedback and e-mail to authors; supporting alerting and prospective sighting functions; delivering content for indexing to secondary publishers and distributors, as well as to Internet indexing services; and supporting individualized access control mechanisms. At the low end, those characterized by PDF-only e-publishing, the cost elements would include simple search, common access control mechanisms, and delivering content for indexing. The range of costs in Internet publishing services is quite wide, although the actual size of this category in the expense budget is relatively small.

The fourth cost category—publishing support—is everything from catering of lunches, to finance offices, including facilities and marketing.

The final category is the cost of reserves. Some organizations have money set aside for disasters or to address opportunities. Some of these reserves are for capital projects or to hedge against key suppliers failing. A great many not-for-profit organizations do not label reserves as such but have investments or bank accounts whose earnings support various programs, but whose principal could be used in a reserve function as needed.

The results of a recent sampling by Michael Keller of six not-for-profit publishers' costs are presented in Table 3-1. The data were compiled in a wide variety of categories, so there is an element of interpretation in these results. The data are presented as a pair of ranges, one from the early 1990s and the other for the most recent data.

---

[6] The cost budgets for *Science* and *Nature*, for instance, would have more elements than specified, and costs for secondary and tertiary publications include different elements than these. Mr. Keller did not cover these elements in STM publishing.

Table 3-1. Sample of six not-for-profit publishers' costs.

Range of Costs of Publication
Publishers n = 5

	1992 or 1993	2002 or 2003
Edit:	28.4%-53.3%	32.4%-56.0%
Print:	25.9%-50.0%	21.8%-27.0%
Mail:	4.3%-16.9%	4.3%-10.0%
Internet:	N/A	4.3%-9.1%
Support:	10.0%-29.4%	3.0%-33.2%
Reserve:	0.0%-12.2%	0.0%-7.2%
Society Money:	0.0%-1.9%	0.0%-2.9%

How might these data be interpreted? First, it appears the publishers have much tighter control now over their budgets than they did 10 years ago. The definitions of categories are better. Second, it is clear that editorial costs have not changed much but that printing, paper, and binding costs are down, at least on a unit basis. Third, the costs of Internet editions have entered the budgets at the level of 4.3 to 9.1 percent.

In addition, respondents to Mr. Keller's inquiry indicated that their publishing budgets have doubled since the early 1990s but that individual subscriptions are almost half in many cases, apparently cannibalized by institutional Internet subscriptions.

Cost increases reported by the not-for-profit publishers have been on the order of 6 percent annually. With a reduction in individual subscriptions, the number of copies of issues printed, bound, and mailed has gone down, but increases in other costs have kept budget numbers from falling.

The reduction in individual subscriptions arises from two pressures. The first is the rising subscription prices themselves or membership dues for individuals. The business models of some publishers involve extra charges for online access. The second is the more easily available content through library servers, which enables a scientist to go to different sources, including especially journals, without having to resort to lots of different user names and passwords. It is simply more convenient to do it that way.

Clearly, the publishers that HighWire has worked with are very concerned about cannibalization and have attempted in many ways to figure out how to cope with it. There may not be a single right answer.

Manuscript submission tracking and refereeing support applications have reduced mailing costs and made it possible for more manuscripts to be processed by existing staff. Increases in the size of the journal in page equivalents, increases in graphics and colors in some instances, and the adoption of advanced Internet features like supplemental information and so forth have increased the costs of Internet publishing services about 50 percent higher per year over other costs.

In short, there is a dynamic balancing act with regard to publishers' costs in the Internet era, with some costs increasing and others decreasing. What is most intriguing, however, is the possibility of removing from 25 to 32 percent of the costs of publishing by switching to electronic journals delivered

8

over the network, and eliminating printing, binding, and mailing paper copy to any subscribers at all.

One might further observe this as a transfer of costs from the publishers to the readers who desire to read articles on paper. Perhaps this is not a new transfer of costs to readers, because many already photocopy articles and presumably cover that time and cost somehow. Eliminating paper editions would also offer some promise of reducing prices to the institutions that are so clearly providing the publishers with the economic basis for publishing at all.

The successful operation of true digital archives—protected repositories for the contents of journals—would permit the removal of the printing, binding, and mailing costs. A true digital archive or repository in the view of most librarians is one that is not merely an aggregation of content accessible to qualified readers or users, but one that preserves and protects the content, features, and functions of the original Internet edition of the deposited journals over many decades and even centuries. True digital archives will have their standards and operational performances publicly known and monitored by publishers, researchers, and librarians alike. Their operations and content then will be audited regularly. However, it is not yet well known how costly automatic data migration will be over time. In any event, It is not enough for a publisher, a library, or an aggregator or other information business to simply declare themselves to be an archive. This function must be proven constantly.

The annual costs of true digital archives can have a broad range. One very inexpensive model is Lots of Copies Keep Stuff Safe (LOCKSS). These are network caches, a design that involves publishers and libraries in voluntary partnerships enabling dozens, maybe hundreds of local caches on cheap magnetic memory, and using very ordinary CPUs. Michael Keller estimates that each of these LOCKSS caches could operate for only tens of thousands of dollars per year.

Another model is that of the large, managed digital repository for multiple data formats and genres of publication. The estimate for operating and maintaining a very large repository, perhaps a petabit or two of data, at Stanford is between $1 million to $1.5 million per year, half for staff and half for technology. HighWire's database now is in excess of 2.5 terabits.

Which institutions will undertake such large managed digital repositories? Almost certainly a few national libraries and university libraries will. But publishers or their Internet service providers could develop and run them as well.

The European Union's laws will soon require deposit of digital editions in one or more national libraries. In the United States, the Library of Congress, along with other federal libraries, promises to develop both its own digital repository, as well as to stimulate and support a distributed network of them. If publishers undertake digital repositories, their costs will enter into the expense budget and, of course, drive journal prices higher.

Another cost currently confronting many publishers is the conversion of back sets of print journals to digital form, providing some level of metadata and word indexing to the contents of each article, and posting and providing access to the back sets. HighWire has done a study on converting the back sets of its journals. They estimate that about 20 million pages could be converted and that the costs of scanning and converting pages to PDF, keying headers, loading data to the HighWire servers, keying references, and linking references could approach $50 million, or about $150,000 per title. Most of that sum is devoted to digitizing companies and other sub-contractors to HighWire Press.

If all this retrospective conversion of back sets occurred in one year, HighWire would have to spend internally about $250,000 in capital costs and about $300,000 in initial staff costs, declining to annual staff expenditure of perhaps $250,000 or $275,000 thereafter. On average, for the 120 publishers paying for services from HighWire that would mean about an additional $2,500 in new operating costs to HighWire Press each year. In other words, the increase in annual costs to publishers for hosting and providing access to the converted back sets would be a fraction of 1 percent of their current expenses each year. These figures, of course, omit any costs for digitizing and other services provided by contractors and subcontractors.

Although the costs of back-set conversion are high, the experience of HighWire suggests that the payoff could be 5-10 times more use of articles in the back sets than is presently experienced. Articles on the HighWire servers are read at the following rates: Within the first three months of issue, about 95 percent of all articles get hits (this presumes that a hit means that somebody is actually reading something). In the next three months, that is, when the articles are four-to-six months old, slightly less than 50 percent of all articles get hits. And when articles are ten months or more old, on average of only 7-10 percent of all articles get hits. However, that rate of hits seems to persist no matter how old the online articles are.

Based on citation analyses, only 10 percent of articles in print back sets older than the online set of digital versions get cited, though not necessarily read. That they should do so is entirely consistent with the belief commonly held since 2001 by publishers associated with HighWire that the version of record of their journals is the online version. This is leading many publishers to digitize the entire run of their titles as the logical next step. In any case, unless other sources of funds are forthcoming, the costs of back-set conversion will become a temporary cost in the expense budgets. Other STM journal publishers, however, have indicated that their back-set conversion and subsequent maintenance costs have been considerably higher.

There are chickens and eggs in expense budgets that also complicate understanding them over time. For instance, the experience of HighWire Press has been that those publishers who first define and design advanced features pay the cost of developing those innovations for those that follow. At the same time, those early adopters reap the benefits of innovation in attracting authors and readers. Eventually, many of the innovative features become generally adopted, and usually at lower cost of adoption than paid by the innovators to innovate. In order to maintain a reputation as innovative, however, one has to continue adopting new features. Some innovation leads to lower costs. For example, HighWire recently announced reductions in prices, thanks to some processing innovations it recently implemented.

Certainly, readers of online journals value the search engines and strategies, the navigation devices, hyperlinks, and availability of various resolutions of images. Alerting services are particularly well thought of too. These sorts of online functions lead users to be more self-sufficient in their search for information and reduce calls upon librarians for help in finding relevant information. Yet it is unlikely that science librarians are volunteering to reduce their staff size as a result.

Concluding Observations

A close study of the costs and benefits of electronic journal publishing from the birth of the World Wide Web would be a good thing to do. It would document, in a neutral way, the profound transformation of an important aspect of the national research effort. That there is likely to be as much change in the next 10 years as in the past decade does not obviate the need for the study. Such a study may help develop new strategies or evolve current ones for accommodating needs of scientists and scholars to report their findings, and for ensuring the long-term survival of the history of science, medicine, and technology.

University budgets are under considerable strain and will be so for at least several more years. Will the deals for access to all journals from a single publisher survive or will there be new deals for multiple titles, cut just right to fit true institutional needs?

Why should there not be a highly diffuse distribution scheme based on authors simply posting their articles on their own sites or on an archive like Cornell arXiv.org e-Print archive (formerly at Los Alamos National Laboratory), and let Google or more specialized search engines bring relevant articles to readers on demand? Who needs all this expensive publishing apparatus anyway?

The answer lies partly in the strong need for peer review of content, expressed variously by most communities of science, and partly in the functions provided by good publishers that are valued and demanded by the scientific community itself. The e-journal survey mentioned earlier showed that the vast majority of respondents want, and have come to expect, a wide array of features making their regular searches for articles relevant to their work easy to find and to use. By implication, the readership focuses its attention in the constantly churning galaxy of new and old articles on a few journals with editorial policies and content that are known and trusted.

Providing relevant, reliable, and consistent levels of content in journals costs money. Highly distributed, diffuse STM publishing with sketchy peer review, dependent upon new search engines to replace the well-articulated scheme of thematic journals and citations in a multidimensional web of related articles, is a descent into information chaos.

Perhaps in the next decade the segment of STM publishing most at risk is the secondary publishers, the abstracters and indexers and the tertiary publishers, those producing the review and prospective articles long after the leading-edge researchers have made use of the most useful articles.

None of the alternative publishing experiments underway or about to get underway operates independently of a larger STM journal publishing establishment, and none operates without costs. Taking the example of the Cornell arXiv in particular, although it is certainly an archive of articles to which many

10

in physics, mathematics, and computer science go to first and constantly, it has had negligible effect, if any, in reducing the number of peer-reviewed articles published in these fields. It may have improved the articles by exposing them in preprint form to many readers, some of whom may have commented back to the author with helpful suggestions. But *Physics Letters* and *Physical Review* have continued to grow, have continued to publish peer-reviewed articles, and those articles have continued to be cited.

It also must be observed that some communities use and read preprints reluctantly. Although there are 200 articles in the *British Medical Journal*'s *Clinical Netprints*, few have received online and public peer review from readers, and fewer, if any, have been cited in that form.

Finally, several experiments in journals that depend nearly entirely on fees paid by authors should be noted. The new journal *Physics*, for example, has published between 25 and 50 articles in each of its first five years of existence. But it has not yet had sufficient citations to be indexed by the *Science Citation Index*.

The change in the business plans of BioMed Central is instructive too. After trying author fees alone, it is selling memberships to institutions so that authors from member institutions do not have to pay for publications. But the membership fee is very high.

The Public Library of Science (PLoS) will enter this list with its first articles in the fall of 2003. It too will depend upon authors' fees for support of its operations. It is starting with an admirable $9 million war chest from the Moore Foundation. PLoS will charge $160 for print copies of its volumes.

The point of mentioning these efforts is that not one has done away with the costs of publishing. Someone always pays. And none of these journals with new business models has become self-sufficient. For the costs of peer-reviewed publishing in science, technology, and medicine to disappear, the requirement for peer review and the demand for thoughtfully gathered, edited, illustrated, and distributed articles must disappear too.

How the experiments in business models might provide competitive pressure on traditional business models and pricing is a topic for discussion and examination over time. Experiments should be tried, but in Michael Keller's view, the solution to the serious crisis of escalating journal costs lies in the not-for-profit societies, whose purposes and fundamental economic model are very closely allied with the purposes and not-for-profit economic models of our research universities and labs.

COMMENTS BY PANEL PARTICIPANTS

Kent Anderson, New England Journal of Medicine

Kent Anderson provided comments from the perspective of the *New England Journal of Medicine* (NEJM), which resides at the interface of scientific research and clinical practice, and which has published continuously for 191 years. The journal mainly serves clinicians, physicians who take care of patients. It is a key translator and interpreter of new science to general and specialist physicians around the world. In many ways, the NEJM situation is unique: It is a large-circulation publication that relies on individual subscriptions, and it is owned by a small state not-for-profit medical society.

Mr. Anderson limited his comments to three areas: how the definition of "publication" is changing; how that change in definition is modifying how the NEJM conducts its business of getting the journal out week to week; and finally, some cautionary notes about conducting a study of publishing costs because of the diversity of STM publications and their users.

The Changing Definition of Publication

Publication used to refer to the act of preparing and issuing the document for public distribution. It could also refer to the act of bringing a document to the public's attention. These definitions served us well for more than 400 years.

Now, publication means much more. It now means a document that is Web-enriched, with links, search capabilities, and potentially other services nested in it. A publication may soon be expected to be

11

maintained in perpetuity by the publisher. A publication now generates usage data.

For publications serving physicians, however, the NEJM has found that physicians are not willing to give up print. They are too pressed for time, and the print copy is too convenient for them. A recent publication by King and Tenopir supported this finding as well.[7]

The Changing Conduct of Publishing at the New England Journal of Medicine

The mission of the NEJM is to work at the interface of biomedical research and clinical practice. As medical and scientific findings become more complex and sophisticated, the journal invests in editors, writers, and illustrators who can analyze and interpret these findings for a clinical audience, so that its readers understand precisely how medicine is evolving.

The NEJM also supports significant costs for Web publishing, data analysis, and reporting. It develops new services at a rapid clip. And it pays for new publishing modalities, such as online-only publication, early release articles, free articles to low-income countries, free research articles after six months, and selected free articles online.

The journal's educational mission has become more complex. It has invested heavily in new continuing medical education initiatives, new ways of illustrating and presenting articles, and new ways of helping users find what they are looking for. It is more important, and in some ways more difficult than ever, to know who the journal's readers are. The staff can no longer look at the print distribution and point to them as its readership. The Web amplifies and concurrently obscures the readership. So, investments in market research, Web data analysis, and warehousing and attending medical meetings with physicians have increased significantly.

Customer service demands have escalated, resulting in the need to develop new software and management systems, the hiring of new staff, and additional phone calls and mail. These service requirements are also adding to the cost of running a publication.

Online peer-review tools in parallel with paper systems have recently added a new layer of cost, with no apparent end to the investment, because periodic software upgrades will be necessary. The skill sets of the journal's expert editors and workers also need continuous improvement to ensure that they can handle all the new inputs and maintain quality.

Marketing the journal's new programs and services is another expense. The word needs to get out that there are new ways to access the journal's information. E-mail systems are part of the marketing and distribution effort now, and these have to be supported, databases need to be developed and maintained, and e-mail notices consistently sent to existing and potential readers.

The pressures to be fast are growing, yet high quality must be maintained. The NEJM publishes information about health, and if it makes mistakes, they can be serious. Recently, the journal published a set of articles on SARS, the Sudden Acute Respiratory Syndrome, and it published them in two weeks or less of receipt—completely peer reviewed, edited, and illustrated papers. They were translated into Chinese within two days of their initial publication and distributed in China in the thousands in print, with the hope they made a major difference. As such demands for faster publication mount, the journal will need to find ways to accomplish them without sacrificing quality, and this is already leading to significant investments in people and systems. Often in these discussions of electronic journal publishing, the people behind all this become obscured.

Issues in Conducting a Study of Journal Publishing Costs

If the publishing costs are to be is studied well, there has to be an acknowledgment of the diversity of the publishing landscape, even in the scientific, technical, and medical publishing area. If only a few publishers participate, the selection bias could drive the study to the wrong answers.

It is necessary to consider which cohorts need to be analyzed. The null hypothesis must be clearly stated. The questions to be asked must be properly framed and a reasonable control group

---

[7] Carol Tenopir and Donald W. King. 2002. "Reading Behavior and Electronic Journals," *Learned Publishing*, 15, 262-265.

selected. In short, a study of this nature could be valuable, but it needs to be well designed, rigorously conducted, and carefully interpreted.

Robert Bovenschulte, American Chemical Society

Electronic publishing has not only revolutionized the publishing industry, it has also tremendously changed the fundamental economics of the STM journal business. Many of these issues overlap both the commercial and the not-for-profit publishers.

What is driving this change? The costs are increasing very rapidly, much more than would have happened if publishers had stayed with print alone. There are two obvious factors responsible for this: the cost of new technology and the increasing volume of publishing.

Costs of New Technology

With all of this expansion of technology costs, the publishers are delivering a much more valuable product to their users. There are enormous new functionalities that are being made available to scientists. The access to information is swift, convenient, and is improving productivity.

Technology now imbues all facets of publishing—from author creation and submission, all the way through to peer review, production and editing, output, and usage. The costs of the technology are not just related to the Web, but apply to all the other technical systems that publishers have to create and integrate. For example, the American Chemical Society (ACS) has 186 editorial offices worldwide for its 31 journals, and all of those offices have to be technically supported. ACS must develop new technologies that support the functioning of those offices and make them more productive.

In 2002, ACS conducted a study through the Seybold Consulting Group with 16 other publishers to assess the future of electronic publishing and, in particular, to get a better understanding of the cost drivers. Both the median and the average cost among these 16 publishers was 5.5 percent going into just the information technology (IT) function. Large publishers allocated less than 2 percent, and the ACS spent about 9 percent. Although it is obvious that larger publishers with a lot more revenue can spend a much smaller percentage on IT, they are nonetheless far outspending the medium-sized and the smaller publishers in the total amount that they can invest in their IT operation. The predictions of the 16 publishers in aggregate were an average of a 21 percent annual increase in their IT spending for the foreseeable future.

The Increasing Volume of Publishing

The ACS has gone from publishing 15,000 articles in 1993, to 23,000 in 2002—a 53 percent increase over that 9-year period. Over the past 2 years or so, the ACS has experienced double-digit increases in submissions. This appears to be largely spurred by the fact that online submission makes submitting an article even easier than before. Moreover, the ACS is receiving a much larger fraction of its submissions from outside the United States.

During this same 9-year period, total costs of publishing at the ACS increased 64 percent, versus the 53 percent increase in articles published. But the cost per article published increased only 7.4 percent. In 1993, for every article the ACS published, it cost $1,712. In 2002, the cost was $1,838. Therefore, there has been quite a significant gain in publishing productivity at ACS, and probably at other publishers, attributable in large measure to technology. Although the technology does cost a lot more, it seems to produce efficiencies.

Digital Repositories

The ACS was one of the first publishers to create a full digital archive, PDF form only, of all of its journals. This entailed a very large upfront cost. In the future, it is likely that preserving the journal literature could be the responsibility of both the library community and the publishers.

13

Michael Keller's notion of trying to move toward one system has a lot of merit. That is not going to be easy to achieve, however. In the interim, the publishers, especially the not-for-profit publishers, feel a very strong obligation to preserve that digital heritage. The end users, and particularly the scientists who write for ACS journals, are not ready to give up print. And, because of the preservation issue, most librarians are not yet ready to give up print either.

Once print versions are eliminated, there might be a savings of 15, 20, or 25 percent of costs. Ending print versions of journals is certainly a worthwhile goal. The position at the ACS is to do nothing to retard the rate at which the community wants to dispense with print. The ACS would be happy to reduce its price increases, possibly even hold them flat, or even provide a small reduction during a period when the print versions are being eliminated as a way of returning those costs to the community. The concern is that in very short order, with the rising volume of publication, the costs of handling those many more articles will in fact wipe out whatever transitory gains there may be from saving on print.

Bernard Rous, Association for Computing Machinery (ACM)

Reasons Why the Costs of Electronic Publishing are Poorly Understood

The costs of electronic publishing are not really well understood at all, and there are some very good reasons why this is the case. First, as has already been mentioned, electronic publishing is not a single activity. One can take PDF files created by an author and mount them on a Web server with a simple index, and that is electronic publishing of a sort. Or one can manage a rigorous online peer-review tracking system, convert multiple submission formats to a single structured-document standard, do some rigorous editing, digitally typeset and compose online page formats, apply style specifications to generate Web displays with rich metadata supporting sophisticated functions, and build links to related works and associated data sets integrated with multimedia presentations and applets that let the user interact and manipulate the data. This too is electronic publishing. It is miles apart and many decimal points away from the first approach.

Second, electronic publishing costs remain fuzzy because we are still living in a bimodal publishing world. Even some direct expenses arguably can be charged to either print or electronic cost centers. Where a publisher puts them often depends on the conceptual model of the publishing enterprise. On the one hand, if the publisher looks at online offerings as an incremental add-on to the print version, then more costs will be charged to the print version. On the other hand, if the publisher considers the print version to be a secondary derivative of a core electronic publishing process, more costs are likely to be charged to the digital side. When it comes to the indirect costs of staff and overhead, the same considerations apply, with perhaps even greater leeway because of the guesswork that is involved in these types of cost allocations.

Third, the decisions to charge costs to a print or to an electronic publication are part of a political process. There are times when you want to isolate and protect an existing and stable print business, so you attribute any and all new costs to the digital side. You may also want to minimize positive margins on the digital side to avoid debate over the pricing of electronic products. At other times, the desire to show that the online baby has taken wings, is self-sustaining, and has a robust future can tilt all debatable charges to the print side. This is not to say that the books are being cooked. It is just the way you look at the business that you are running.

Fourth, it is very difficult to compare print and electronic costs, because the products themselves are not the same. The traditional average cost per printed copy produced is meaningless and very hard to compare to the costs of building, maintaining, operating, and evolving a digital resource as a single facility.

Fifth, accounting systems sometimes evolve more slowly than shifts in publishing process. New costs appropriate to online publications are sometimes dumped into pre-existing print line items.

Sixth, electronic publishing has not reached a steady state by any means. There is still lots of development going on, some of which lowers costs, and some of which raises them.

These are some of the reasons why the costs of electronic publishing remain somewhat obscure and also why a study of those costs would be both very difficult to carry out and very important to attempt.

14

Unanticipated Costs of Electronic Publishing

It is also instructive to mention several components of electronic publishing costs that were not fully anticipated by ACM when they went online. First of all, customer support costs have been phenomenally different from the print paradigm. Not only is there a larger volume and variety of customer complaints and requests for change, but the level of knowledge and the expertise required to answer them is much more expensive.

The cost of sales is higher. The product is different, and the market is shifting. ACM no longer sells title subscriptions but rather licenses access to a digital resource. The high price tag for a global corporate license or for a large consortium means that more personal contact is required to make the sale. Furthermore, such licenses are not simply sold. They are negotiated, sometimes with governments, and this requires much more expensive sales personnel.

Digital services are also built on top of good metadata, and metadata costs are high. The richer the metadata, the higher the costs.

Subject classification is costly as well. The application of taxonomies is a powerful tool in organizing online knowledge. The costs for building what has been referred to as the "semantic Web" are still largely unknown. There are a surprising number of opportunity costs. There are so many new features, services, and ways of visualizing data and communicating knowledge. There is a lot more work that can be done than has been done so far.

Finally, some upfront, one-time investments in electronic publishing turn out to be recurring costs, and some recur with alarming frequency. For example, ACM is in the middle of its fourth digital library interface release since 1997.

Gordon Tibbitts, Blackwell Publishing, USA

Is there a cost-tipping point in e-journals, that is, the place where all of a sudden the costs will dramatically drop, and there will be a new golden era of electronic publishing?

There are two classes of commercial publishers that need to be considered in this context. One category of publishers provides publishing services for professional societies. The other type of publisher owns most of the actual material they publish. Figure 3-1 compares the current print plus electronic publishing costs for these two types of for-profit publishers. The major difference between the two is that society-oriented publishers pay a lot more royalties—profit shares, royalties, stipends—back to the societies, who are the gatekeepers for that peer-reviewed information. The other publishers keep most of their profits.

15

For-Profit Heavily Society Publisher   For-Profit Mostly Owned Publisher
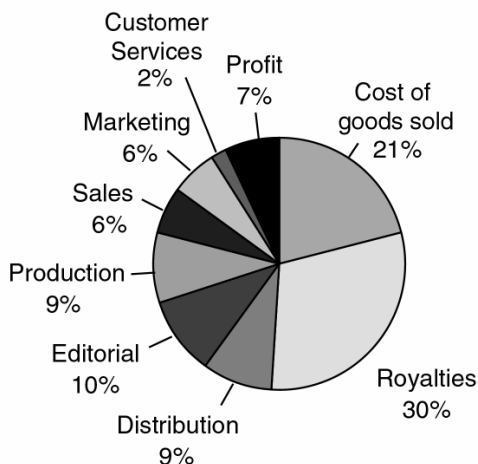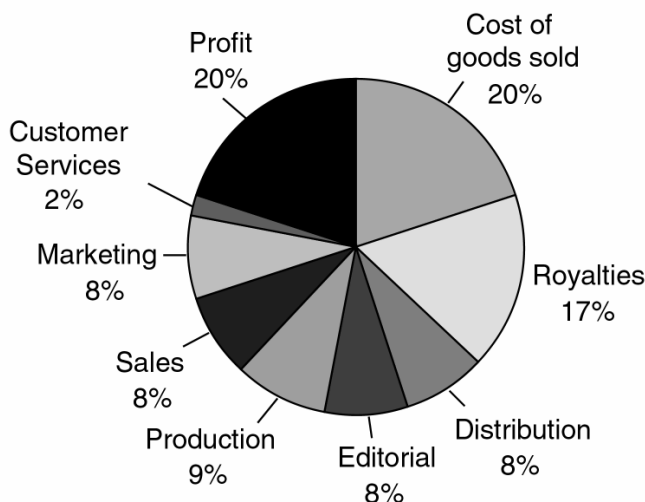
FIGURE 3-1. Current Print + e-Costs of For-Profit Publishers (% of Sale)

There is another point worth mentioning, an issue of scale. In the marketing and sales statistics in Figure 3-1, there is about a four point difference between the two types of publishers, but the mostly owned publishers have billions of dollars in revenues, so those four percentage points translate into massive marketing dollars.

Figure 3-2 presents the current electronic journal cost components. The pie chart on the left looks at "e-incremental costs," or those things that could be separated out as being purely new costs that the business is incurring that are electronic. Many of the societies have a fundamental bylaw in their charter to disseminate information, sometimes for free. Nonetheless, there is a cost to that, which is quantified under "marketing" at about 7 percent.

16

e-Incremental Costs

e-IT Costs as a % of O/H

* IT is 16% of overhead
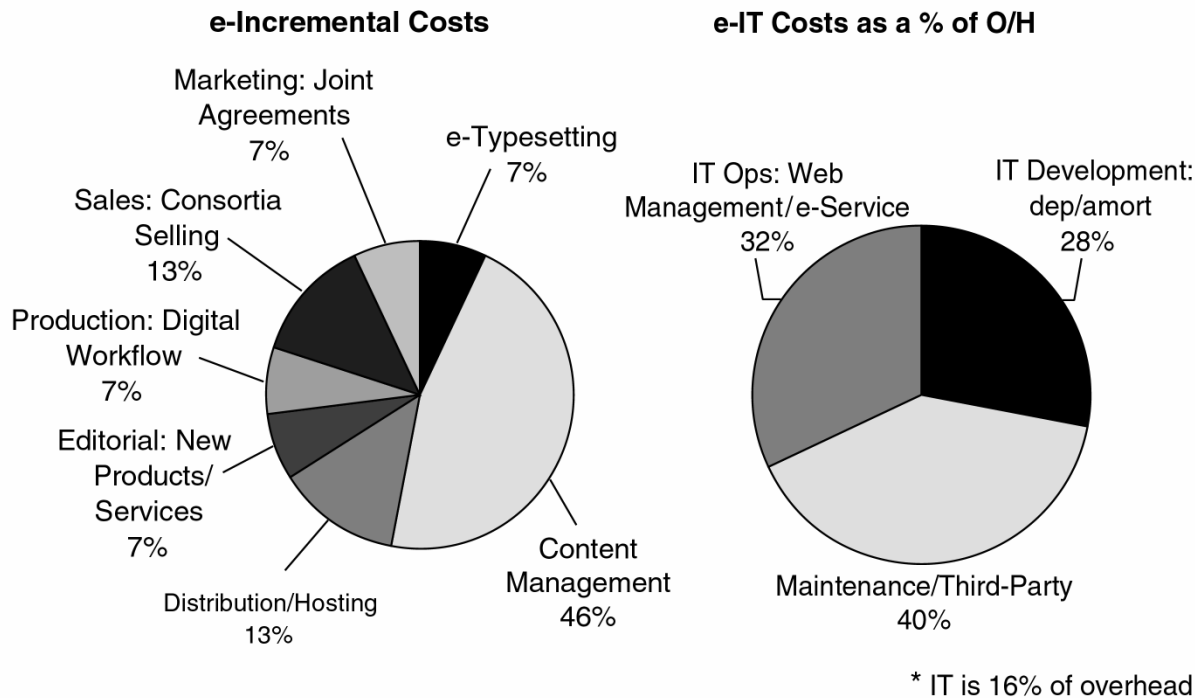
FIGURE 3-2. Where is the cost tipping point?

There is consortia selling, a brand-new thing. Publishers have to negotiate with governments and large institutions, and that costs quite a bit of money—perhaps 13 percent of the e-incremental costs—to have large consortia selling for the journal.

The biggest cost in Figure 3-2 is content management. This includes many people whose job is to get the information, organize it, and disseminate it. This new dissemination can be through Web sites, Palm Pilots, libraries, and online systems. Of course, this is not new; this is the information business. In an information business there are three primary costs for developing and maintaining information and for operations. Every one of the new e-incremental costs has those components. You do not build something once. You go through versions 1, 2, 3, 4, and 10. Microsoft knows that very well; they keep upgrading their versions. There is an ongoing requirement to pay licenses, and ultimately humans are needed to actually operate these systems.

Figure 3-3 provides some diagrams of relative print (P) and electronic journal (e) costs. The Venn diagram on the far left shows there is an intersection between P and e costs. There are some costs that are purely print or electronic. Perhaps, since there is a lot of intersection of costs, there will be a collapsing of the cost structure of the electronic journal, resulting in a small p (small cost in print) and a small e (a small cost in electronic).
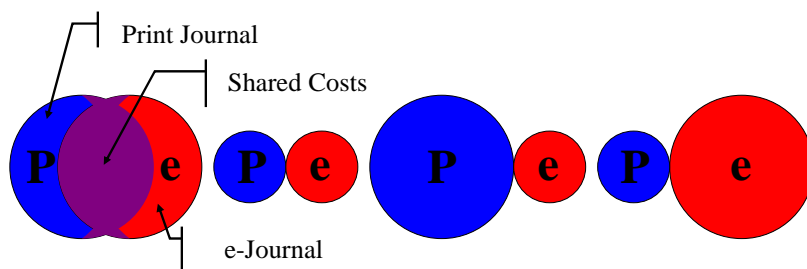
FIGURE 3-3. Schematic Diagrams of Scenarios of Relative Print (P) and Electronic (e) Journal Costs

More likely, the current situation has a large print cost and a small electronic cost. Or, as Bernard Rous put it, you could play accounting games, and you could actually go with a little p and a big E depending on what you are trying to accomplish on any particular day. Unfortunately, Gordon Tibbitts thinks that a big P and a big E are the most likely outcome.

Some trends—both positive and negative—affecting cost include the number of articles, supplementary data, back issues, customer support, new on-line features, speed requirements, archive/repository alternatives, and new business entrants. Technology trends are driving costs down, for example, Microsoft Office 03', Word 11, and PDF+. Trends are down in individual subscribers in favor of consortia, the amount of money everyone has to spend, and perhaps in flagrant pricing. There are rapidly rising institutional prices responding to the shift to online publishing, the drops in individual subscribers, and in "acknowledged" business risks. The complexity of layout, graphics, and linking increase cost and the need for speed, as do the breadth of content and quality requirements.

Rather than a cost-tipping point, there is a cost-shifting point. It is shifting away from print toward electronic. We will see a lessening in the big P over time, but a big E will take its place. There will not be any savings.

If we are to achieve any cost savings, we have to stop trying to invent the moon landing. We do not need very complex systems. Some of the information just needs to get disseminated. Some technologies are simpler and more widespread, and we probably should embrace them to keep costs in check. Reducing the number of new versions of the same basic application is another way to control costs.

Finally, and most importantly, when many publishers get together and share standards and archiving, and collaborate, that is certainly a very good way to reduce costs across the board.

18

DISCUSSION OF ISSUES

What Would a Study of Journal Cost Accomplish?

Robert Bovenschulte began the discussion by noting that doing studies of electronic publishing costs has major definitional and methodological challenges, although that is not to say they cannot be done. If such a study were done perfectly with perfect information, what would one then do with the study?

Michael Keller said that the answer would depend on the position of the user. As a chief librarian responsible for managing a fairly substantial acquisitions budget and for serving many different disciplines, he would hope that such a study would affect the responses of the library community as a whole to these problems. The responses so far have been so fragmentary that librarians have not influenced the situation as effectively as they might.

From the perspective of an Internet service provider, that kind of study would help in the understanding of the comparative benefits or detriments of the various methods and operations, and cause Internet service providers to become more critical in some ways than perhaps they are at the present time.

As a publisher of books for Stanford University Press, Mr. Keller hopes constantly for some magic bullet that would help him take control of those costs, to better attain a self-sustaining posture. He thinks, furthermore, that it is very important for the research community to have a common understanding of what the problem set is. He would expect them, therefore, to elevate the level of discourse to something a lot more scientific than it has been to date.

Moving from Print to Electronic Versions

Bernard Rous noted that in several of the panel comments, it was implied that if only the publishers could get rid of print, there would be a huge savings in the publishing system. Although there is some truth in that, it is also true that although the publications of the ACM are rather inexpensive, there is a margin in the print publishing side of the business that the society cannot afford to do without. As such, there is not only a demand for print versions from some customers, libraries, and some users, but also from the publishers themselves, since the profit margins that are realized from the print side are actually necessary for them to continue in business.

Michael Keller responded that in observing the information-seeking and -using behaviors of his 15-year old daughter and her cohort, he has seen an almost total ability to ignore print resources. Although they certainly read novels and sometimes popular magazines, when they are writing a paper, most of them use the Internet entirely. Regular features of these papers now include media clips, sounds, and various graphics.

It will only be 10 years before that cohort is in graduate school. Another five years after that, some of them will be assistant professors heading for tenure. The students at Stanford are already driving the faculty in the same direction, and a great many of the faculty have figured it out as well.

Mr. Keller agreed completely with Kent Anderson's point that, for a variety of reasons, there are hosts of readers in some communities for whom online reading and searching are not presently good options. Yet for many research communities, especially in the basic sciences that advance rapidly—such as physics, chemistry, mathematics, biology, and geology—there is real promise for removing the print version of journals altogether. That does not necessarily mean, however, that there will be a total removal of those paper, printing, and binding costs.

Robert Bovenschulte noted that the ACS adopted a policy five years ago of trying to encourage its members to stop subscribing to individual print journals, and instead to obtain them online from their institution's subscription. The ACS creates a significant pricing differential between print and Web subscriptions, the electronic ones obviously being much cheaper. About three years ago the ACS created a new alternative model to print plus Web cost, which was Web cost only, with print being a low-cost option.

This approach was pioneered by Academic Press. This model has a lot of merit, because it makes publishers think about what the true costs are and what margin they need to sustain the business

19

economically, and not get caught up with the fact that the reporting system tends to indicate that the print version is profitable.

Michael Keller added that the American Society for Biochemistry and Molecular Biology (ASBMB), when it first came out with the *Journal of Biological Chemistry* (JBC) online, selected two separate models, one based on the price for the print subscription, and a separate lower price for the electronic-only version on the grounds that they did not have to print on paper. They were trying to encourage institutional subscribers in particular to migrate over to that electronic-only version.

Floyd Bloom next raised a question from a Web listener, who asked, Why if the *British Medical Journal* (BMJ) does everything that the *New England Journal of Medicine* (NEJM) does, and is available to the public without charge, can't the NEJM be made similarly available?

Kent Anderson responded that the BMJ sees the future in a different way than the NEJM. The NEJM provides a lot of free access for people who truly cannot afford it, people in 130 developing countries, and its research articles are free upon registration after six months. Nevertheless, he thinks the BMJ experiment is worth watching. They have a national medical society that has the means of supporting their publishing efforts, so they are able to conduct this experiment. The NEJM has chosen a path that relies on knowing that it is publishing information that people find valuable enough to pay for.

The Massachusetts Medical Society has 17,000 members in Massachusetts, but the circulation of the journal in print is between 230,000 and 240,000. The exact readership is hard to estimate, however. Mr. Anderson thinks it is somewhere between 500,000 and 1 million readers per week.

Dr. Bloom added that the British Medical Association, the publisher of the BMJ, has a required membership for anyone who wishes to practice medicine in the United Kingdom. Therefore, they have a huge subscriber base whether they sell it or not. They also have other publications for which they charge.

Digital Archiving Issues

Floyd Bloom asked Mike Keller if the true digital archive was achieved and allowed us to diminish the reliance on print, what organization would be responsible for trying to come to some agreement as to what the taxonomy and terminology should be for that kind of archiving?

Mr. Keller responded by reiterating that the support for the publishing industry does not come from individual subscribers, who receive their copies, whether online or in print, basically at the marginal cost of providing them. The real support for the publishing industry comes from institutional subscriptions from libraries and labs. It is the libraries and labs that are most unwilling at this time to give up paper, because no one has a reliable way to store bits and bytes over many years and over many different changes of operating systems, applications, data formats, and the like. When this problem is resolved, the libraries will be challenging the publishing industry to get rid of paper, or at least not deliver those costs to them.

So, who is going to do it? There are efforts underway in Europe. For the past couple of years, the Library of Congress has managed a big planning program, the National Digital Information Infrastructure and Preservation Program, with sustained congressional funding to support several different experiments around the country. There also is funding from the National Science Foundation (NSF) for the digital library initiative, and some of that money has gone to projects like LOCKSS. Finally, of course, there are industry segments that are quite interested in this. There are suppliers of content management systems, metadata management systems, and magnetic memory anxious to get involved.

In the next year, there will be some initial efforts to demonstrate the viability of these. We are going to have to simulate the passage of time and of operating systems and so forth, but the reliability of digital archiving can be improved. Within five years we will see general acceptance of some methods. Mr. Keller does not think that there will be a single system, but rather an articulated system with many different approaches.

Bruce McHenry, Discussion Systems, noted that all of the panel members have publications where peer review is the crucial distinction from simply having authors publish to the Web themselves. He asked each panelist to describe the unique parts of their process, the best and the worst things about it, and how they would like to see it improved.

Kent Anderson said that the peer-review process at the *New England Journal of Medicine* is rigorous. First, internal editors review paper submissions as they come in and judge them for interest, novelty, and completeness. If they move on from there, they go out to two to six external peer reviewers.

When those reviews come back, they are used to judge whether the paper will move forward from there. If it does, it is brought before a panel of associate editors, deputy editors, and the senior editors for discussion, where the paper is explained, questions are asked, and it is judged by that group, usually during a very rigorous discussion.

Then it is decided whether it moves on or not. If it does move on, it goes through a statistical review and a technical review. The queries are brought to the author, who must answer them or the paper does not move forward. The peer-review process lasts anywhere from a few weeks to a few years, depending on the requirements. Sometimes the NEJM asks the authors to either complete experiments or to give additional data.

As far as weaknesses in the peer-review process, or ways it could be improved, Mr. Anderson thinks that one of the concerns now is that the time pressures in medicine are so great that finding willing peer reviewers is increasingly difficult. The situation could be improved by having the academic community understand the value of this interaction in the scientific and medical publishing process, and by having some sort of reflection of that in academia as a reward system so that people continue to devote the time they need to the peer-review process.

Dr. Bloom added that as the number of submissions has risen, the number of people available to provide dependable reviews of those articles has not increased. So publishers are calling upon the same people time and time again to provide this difficult but largely unpaid service.

Robert Bovenschulte said there is a lot of commonality at the ACS with what Kent Anderson described, except that ACS publishes 31 journals. The editor-in-chief of each journal has some number, in some cases a large number, of associate editors. These are typically academicians scattered around the world in 186 offices. They basically handle anywhere from 250 to 500 manuscripts a year. This is an incredible load, and they do not have the second level that Kent Anderson described, where the editors meet and decide. That is done by an associate editor under the general guidance and policies of the editor-in-chief. The rejection rate is about 35 percent, although it varies from journal to journal. As more articles are submitted, there is additional work and costs. New associate editors and new offices need to be added in order to handle that load, even if the rejection rate goes up.

Finally, there is agreement with Kent Anderson's observation that it is harder and harder to persuade people to serve as editors-in-chief, associate editors, reviewers, or members of editorial advisory boards. They simply do not have the time. The best people are in enormous demand. The Web helps and has the benefit of speeding everything up. But there is the potential for a meltdown of the system in the next five years or so, particularly for those journals that are not seen as quite as essential to the community as others.

Bernard Rous next pointed out that the ACM publishes several genres of technical material, and each receives a different type of peer review. The review of technical newsletters is generally done by the editor-in-chief.

Conference proceedings in computing are extremely important, and there is a very different review process for them. There is usually a program committee, and each paper is accepted or rejected upon review by any number of people on that committee, depending on the conference. The rejection rates for certain high-profile conferences are even higher than in some of the equivalent journals in the same subject area. There is no author revision cycle in the peer-review process for conference proceedings, and there is no editorial task.

The ACM also publishes magazines, where the material is actually solicited. There are very few submissions that come in over the transom, so there is a different review process there, and the emphasis is not on originality of content. The material is being written for very wide audiences, so there are different editorial standards. There is a heavy rewriting of these articles in order to achieve that, and that can be very expensive.

Finally, the ACM has about 25 journals that follow the typical journal review process, rigorous and independent review, generally using reviewers, with an author revision cycle and edit pass.

Gordon Tibbitts noted that Blackwell publishes 648 journals and the primary component of all of them is they are peer reviewed. Blackwell interrelates the societies because it has so many; it constantly commingles boards and introduces the editors of journals to one another, so that there is a good cross-section of board membership, not just nationally but all around the world. That is an added value that Blackwell brings to the process.

Blackwell also has spent considerable effort in automating the electronic office. The publisher has partnered with several other firms that work with electronic editorial office management systems,

which has really facilitated a better, faster peer-review process.

The final important characteristic is that Blackwell pays its editors. The publisher has been pressured quite heavily for many years now, and it has given into that pressure. There is a need to pay editorial stipends to the people who are working hard on the journals. It does drive profits down, but it also retains the editors. Many of these people have five or six jobs already—as chairs of medical institutions, authors, practicing surgeons, and professors—and it provides them with some extra funding so that they can hire support staff and things like that.

e-Books Different from e-Journals

Dr. Mohamad Al-Ubaydli, from the National Center for Biotechnology Information (NCBI), asked why electronic books are more expensive than books in print if electronic journals are cheaper to produce than the print version. Is there a difference in the cost or is there a marketing issue?

Gordon Tibbitts responded that although journals and books are similar, they are different product lines. The unit cost in books, and elasticity of that demand, determines the pricing. There are more people sharing e-books freely, even though they should not be. The past three years have seen eight bankruptcies in large organizations that have tried to make money publishing books online. So, it is a very dangerous business area, and people do charge more for the online versions because they realize they will get fewer unit sales. Michael Keller added that customers might be charged more for an electronic book because it can do more things, such as better searching, moving pictures, or other functions.

There are other business models for e-books, however, that include reading and searching for free, paying on a page rate per download or printing, or paying a monthly charge of $20 or so to get access to tens of thousands of books. It is an industry in development and different from the science journal business.

Difficulties in the Transition from Paper to e-publishing

John Gardenier, a retired statistician from the National Institutes of Health (NIH), commented that the electronic medium is going to require a major reconceptualization of the entire process of demand, supply, process, technology, and so forth. One of the most significant changes has been the transition from seeking copyright protection for the expression of facts and ideas, to seeking protection and ownership of the underlying facts themselves. This became a very serious issue in 1996, with the enactment in the European Union of the directive on the legal protection of databases. It has been a continuing battle ever since in the United States.

If the flow of scientific information is restricted either by cost or by legalities, the capability of a society to continue to generate technological innovation diminishes, and, therefore, it becomes disadvantaged relative to what it was before, and possibly to other societies. The reason this pertains to this session is that it is going to take new sources of money and new sources of ideas in order for the not-for-profits at least, and perhaps the for-profits as well, to complete the transition to electronic publishing, while maintaining academic freedom and innovation capability.

A lot of that money can come from sources that have not yet begun to be tapped. The NSF was mentioned but industry has a huge stake in this as well, not just manufacturing but service industries, information industries, and so forth. We, as a community, need to do a better job of outreach to industry, and having them help us to resolve these problems.

Michael Keller responded with two points. First, the STM publishing community in general has gone a long way in its transition to electronic publishing. Second, a stable position cannot be achieved. There will continue to be changes in delivery, in possibilities for new kinds of reports, and in distribution that will mean a great deal to the scholarly community. These publishers already have been quite innovative and have moved a long way from where they were 10 or 20 years ago.

Cost-Containment Strategies

Fred Friend, representing the Joint Information Systems Committee in the United Kingdom, said there was a consistent message from the publishers of the panel that their costs are increasing. What was not discussed was any strategy for dealing with that situation apart from just a few positive comments by Gordon Tibbitts. If the only strategy is to increase the subscriptions paid for by the libraries and laboratories, then there is no future in that. If the publishers are to survive, they must have a strategy for coping with increasing costs.

Bernard Rous responded that the panelists may have created a misimpression since they were addressing costs somewhat in isolation. Costs need to be taken into account together with access and with pricing. He believes that the cost per person of accessing the body of research has plummeted dramatically in the electronic context, even with the increasing costs.

Robert Bovenschulte added that the system is the solution and that we are beginning to see real economies now. He documented some of these in his opening remarks, such as using technology to make staff more productive or dealing with fewer staff to do an increased amount of work. Unless publishers continue to drive the technology to help them be more efficient, however, they will not get out of the box in which they are in danger of being trapped.

Gordon Tibbitts thought that publishers probably should not be in the archiving business. Libraries are probably better positioned to do that. Publishers are innovating in ways that certainly are creative and are making them money, but in some cases their innovations probably should be shared, rather than remaining proprietary.

There are many expenses publishers incur because they are still thinking in terms of printed copy and subscriptions, and not thinking of products. When publishers realize that their value-added might be in the branding of the society, their assistance to the peer-review process, and the move away from some of these technological innovations, the costs will go down, and they can then charge less. The STM publishers are still in the infancy of learning who they are.

There are many different ways of collaborating that can bring costs down dramatically. For example, open access and shareware of various types of search engines and online systems. There is no reason why publishers have to have proprietary online systems. Those are the kinds of things that if shared collectively—with funds from grant organizations, governments, publishers, and societies—can build a better environment.

Publishers do have value. They have value in their ability to facilitate the scientific process. They are collectively connected to more of the distribution channels, so there are economies of scale. And they add a lot of value in funding and innovating. Blackwell and other publishers invest in projects that do not have a payback for 15 years, or perhaps ever. New societies are born, new science is funded, and that is something publishers add to the community.

Michael Keller said that his solution is very different. Libraries should not necessarily support all scholarly communication efforts that are brought to them. He thinks it is irresponsible for libraries to subscribe to the big package deals in which fully 55 percent of the hits in those deals come from 15 percent of the titles. The small societies that want to publish should do so and should be supported in that. The incredibly elaborate overhead that a lot of publishing now entails is unnecessary.

The Cornell arXiv and similar efforts provide one possible solution to this problem. It would be fine if 50 percent of all the scientific journals that are published today disappeared. Those articles that have been through a cascade of rejections from journals could go into preprint archives, and receive recognition later by the number of times they are cited. Reducing the flow of formal publications in half would be the goal.

Kent Anderson observed that there is some risk involved in this. It is difficult for a publisher to know whether it is adding enough value to make a difference so that the investments will pay off. There may be some strategies for containing these new costs and bringing them in line, but at the NEJM their hope is that they are adding enough value so that, over time, the small risks that they are taking in multiple channels will pay off for them.

Vulnerability of Secondary and Tertiary Publishers

Jill O'Neill, with the National Federation of Abstracting and Indexing Services, noted that during the course of Michael Keller's presentation, he suggested that secondary and tertiary publishers are perhaps in a more vulnerable position just by virtue of what they do. Yet, at the same time, he expressed his concern about the next generation of scientists not being well enough equipped to know where to go for research resources that are outside the Internet publishing environment. Is the problem that the secondary services do not add enough value, in which case they have to focus more on information competency to teach the younger researchers how to do those appropriate searches, or do they somehow need to develop better mechanisms for searching and retrieving the precise answer from this huge corpus of literature?

Michael Keller responded that he is supportive of secondary and tertiary publishing and has himself contributed to secondary publishing. He believes that the secondary publishers in particular are vulnerable, because they are being overtaken by the broad general search engines and by the assembling of peer-to-peer kinds of understandings about what is being used—the sort of feature and function you get when you go to Amazon.com to buy a book. You select the book you want to buy, put it in your cart, and go to check out, where they inform you that there are six more books that others who have bought this book have found very interesting. That sort of function could overtake the secondary publishers.

The secondary publishers have to find some ways of being more effective, more precise, but also more general. They have to look for many different approaches, and some of this must be done automatically. However, if we end up with a highly distributed and diffuse situation in which authors place their contributions on individual servers, then there will be a huge role for secondary publishing.

Bernard Rous added that he thinks the secondary publishers, as they are now, are vulnerable, but not because secondary publishing is vulnerable. What is currently happening is that primary publishers, in order to go online, have to create secondary services. They are in the business of collecting all of the metadata that they can, and presenting those as a secondary service that lies on top of full-text archives. Secondary publishing is merging with primary publishing, especially in the aggregator business.

Electronic Publishing by Small and Mid-sized Societies

Ed Barnas, journals manager for Cambridge University Press, noted that the panel is representative basically of big publishers, and there are certain economies of scale in production and technologies in developing their systems. He asked the panelists to address the question of costs in converting from print to electronic by the mid-sized societies that publish their own journals, or by mid-sized publishers, because that is not an inconsiderable concern on their part.

Gordon Tibbitts responded first that with innovations from commercial vendors, such as Adobe and Microsoft, and simple services, such as e-typesetting in China, India, and other places, the costs for both small and large publishers alike will be lower. He also advocated the use of simpler, standard formats that all publishers agree to adhere to, and removing the barriers of very complex Web sites, which are out of the reach of even publishers the size of Blackwell.

Kent Anderson offered the thought, just to be provocative, that it may be even riskier for a mid-sized society to stop delivering print. The situation there may be harder to change, because it is perceived by members to be such a strong benefit.

Having completed two terms as chair of the publications committee for the Society for Neuroscience, a 30,000-member society, Floyd Bloom noted that, in 1997, the decision was made by the society's president that online publishing was the way to go, and they were one of the early journals to join the HighWire group. Within 2 years the individual subscribing base went down by 93 percent, but the readership went up nearly 250 percent, and members all over the world were thrilled to be able to have their information just as soon as those members residing in the United States had theirs. The society recognized, however, that the cost of doing so was considerably more than an initial one-time cost. Therefore, a segment of the dues, at least for the present time, has been earmarked specifically for the

24

support of the electronic conversion. Dr. Bloom then asked if there was anyone in the audience who wanted to comment for a 3,000-to-10,000-member-size society.

Alan Kraut, with the American Psychological Society (a 13,000-member society, but a new organization, only 10 years old), said that they published a couple of journals. Over the course of their brief lifetime they have several times tested in quite elaborate detail whether they could publish their own journals online, and recently have decided they cannot. The risks, particularly for a small organization with modest reserves, simply cannot be taken. They remain a very satisfied partner with Blackwell publishing.

Martin Frank, with the American Physiological Society (about 11,000 members), noted that they publish 14 scientific journals. None of those circulations comes close to what the *New England Journal of Medicine* does. The society has been publishing electronically for about 10 years, beginning on a gopher server back in 1993.

The reason small societies like the American Physiological Society can manage to publish electronically is because of the compounding effects of information exchange that occurs at the meetings that they have with HighWire Press. Technologically, he does not think the society could have done it alone. The fact that there are partners in the nonprofit STM publishing sector that come together with HighWire has made it possible for them to publish all their journals electronically.

Donald King, of the University of Pittsburgh, commented that if you look at the overall publishing cost, and divide it by the number of scientists, the publishing costs seem to be going down. If you look at it from the standpoint of the amount of reading, it is going down even more, because scientists seem to be reading more on a per-person basis. One of the concerns he has is that publishing has traditionally been a business in which some portions of each publishing enterprise subsidize other portions. That is, in any given journal there will be some articles that are not read very much. In other cases, there may be very high quality journals that only have a small audience. And, within a publisher's suite of journals, some of those journals are being subsidized by other journals. When looking at the future system, there needs to be a recognition that there may only be a small audience for most articles, but the fixed costs for all articles are going to be the same. There is a need to be able to continue to satisfy the demand for all of them.

The Costs of Technological Enhancements

Pat Molholt, Columbia University, noted that all of the panelists have said that publisher IT costs are rising, and she questioned whether they have to do that across the board. Why not have both a vanilla version that may in fact satisfy many people, as well as an enhanced version for those people who need an animated graphic or some other added functions and for which they would pay some additional fee? The technology then can indeed be developed collaboratively, so that there is a standard plug-in when that is needed, and if it becomes important to the article being read, it can be invoked, but it would not have to be paid for all the time.

Robert Bovenschulte responded that this is an excellent point, and it is one of the questions that the ACS has wrestled with over the past four or five years in its strategic planning—questions such as how many innovations are needed and what costs should be introduced. He has seen some anecdotal research suggesting that what the scientist who is reading really values is the search capability, rapid access, and linking. Regarding the linking function through cross-referencing, there is now an industrywide, precompetitive endeavor that has great value.

The various technological enhancements can have value, however, for individual communities and for editors. There is pressure on the publishers to do something to compete with what other publishers have done. The publishers are competing less for the library dollars than they are for the best authors. Of course, that applies to editors and reviewers as well.

On the other hand, many of the recent innovations that tend to drive up the IT costs are used very little and are not of great value. However, it is very hard to predict which innovations will prove valuable in advance. Moreover, even an innovation that is not much used today may turn out to be one that is very valuable 5 or 10 years from now. So, the publishers are all experimenting and trying to stay on the learning curve. If you get off the learning curve as a publisher, you are in deep trouble competitively.

Kent Anderson added that the NEJM consistently provides new technological upgrades or services in response to its readers' comments, and, in some cases, based on a little bit of guesswork.

Although some of these new technological capabilities are not highly used, more often than not they take off and have thousands of users. Then the journal has to support the application, iterate and improve it, and find out what it indicates about what other needs the customers may have. He has been surprised at how much demand there has been for new applications.

Michael Keller pointed out that when hyperlinking began in 1995, no one had done it before, and now it has become an absolute standard. The same is true about inserting Java scripts showing the operation of ribosomes with commonly available plug-ins. People began to use them. Some of these technologies that are developed on the margin eventually become quite popular.

Bernard Rous said that it is difficult to decide which innovations to pursue and which to leave, but the suggestion that there be differential service levels also presents problems. The ACM wanted to maintain a benefit for membership in the organization, as opposed to institutional subscriptions, so the society started differentiating levels of service, depending on whether someone was a member or a patron of an institution that licensed an ACM journal. But differentiating service levels adds real cost and a lot of complexity to the system. It also adds a lot of complexity for the end user, and a lot of trouble for the society to explain the differences to the different users.

Finally, Lenne Miller from the Endocrine Society pointed out that while most of the participants in this session are publishers, they all have different business circumstances. For example, the Endocrine Society has four journals and 10,000 members. The society is rather tied to its print clinical journal because of the revenue generated by the $2 million per year of pharmaceutical advertising in it. However, the Endocrine Society does not get nearly as much revenue from its annual meeting as, for example, the Society for Neuroscience does. So, there are different business models and different business circumstances and these must be kept in mind.

4. PUBLICATION BUSINESS MODELS AND REVENUE

INTRODUCTORY COMMENTS

Jeffrey MacKie-Mason, University of Michigan

Scholarly publishing in any medium requires substantial resources beyond the content creation costs. In the early 1990s, when digital publishing started to become a serious possibility, there was some confusion about this in some corners of the scientific community. There was a sense that publishers did not really do anything. Everything was done by authors, and publishers were valueless middle people who distributed the material to readers. As soon as the Web emerged in 1994, it began to be thought that the publishers could disappear.

That, of course, simply is not true. Publishers are providers of value-added products and services. They perform some very crucial functions by producing documents in different media, doing the editorial and design work, marketing the material, and getting readers connected to writers, and so forth. All of those functions involve costs. Even a not-for-profit publisher has to recover its costs and have some sort of reserve. The for-profit firms need to get a return on their investment.

This session, therefore, is devoted to a discussion of sources and types of revenue, ways of raising revenue, and different business models, particularly in a world where digital publishing is becoming much more the norm. The defining question is: How does a publisher organize delivery and rights management across modes of access so as to recover production costs and induce ongoing investment in development? As discussed in the first panel, new technologies are emerging all the time; new forms of delivery, new forms of access, and new types of services are being provided. That requires an ongoing stream of investment for things that once were thought of as one-time costs.

What are the sources of value? Revenue is simply a transfer of value between parties. To address a business model, particularly to start thinking outside the box and think about changing frameworks, we need to think about who values the publications or the information delivery in the first place. Who might be willing to share some of that value back to the publishers and the distributors?—the teachers, public and private researchers, students, practitioners, general public, and so forth.

Another question is, Where does the publisher collect that value? What is the best way to organize the extraction of value from some of the stakeholders and transfer it to some of the providers and distributors? Is it through the universities?

We heard that the traditional publishers are focusing increasingly on library subscription models as their primary mechanism of collecting revenues. In some cases it is from the practitioners, for instance the medical practitioners who subscribe to the *New England Journal of Medicine.* There has also been much discussion about the delivery of scientific materials to developing countries, and a question about whether or not those should be delivered free as part of an expansion of access to knowledge around the world or whether developing countries should be contributing as well in some fashion.

Who are the providers and who are the potential toll collectors? There are a number of different services that go into the publishing process, and the publisher does not necessarily provide all of them. Are those services going to be provided as a bundle by a single service provider, or will there be multiple providers providing overlapping services who will need to recover their costs, and provide different aspects of access to information content?

The business models depend on how the information is accessed and what the information is. Again, in a digital world, we no longer need a single standard mode, a journal. We can think about breaking up the information in many different ways, and repackaging and distributing it in different combinations. This often is referred to as the unbundling and rebundling phenomenon. The corpus of information can be broken apart into its constituent pieces, particularly if it has been nicely organized, with a good markup language and tagging system, and then repackaged in other ways. So, we can sell it by

27

the view, by multiple views, or by time periods of viewing, among many other approaches.

There also can be charges for printing and file usage, as opposed to merely viewing usage. You could allow people to keep local electronic copies for rapid access and local archiving, or there could only be server copies. It is a question of whether you can search the full text or the image only. Bernard Rous mentioned that this is an issue that the ACM has faced as a membership organization. Which services should it provide to the general public, and which enhanced services should it provide to its members in order to induce them to remain members?

What is the role for government revenue? Much of what we are talking about—scientific, technical, and medical information and scholarly research—is information that benefits the general public either directly or indirectly, far beyond the community of scientists and scholars who are using it. There is a public interest in the dissemination of knowledge, in addition to its creation. Of course, governments in most countries are one of the main, if not the main, sources of funding to create the knowledge in the first place—to fund the research. With some exceptions, however, they have not been one of the main providers or direct disseminators of that information. Ultimately, at some level, governments are paying for it through subsidies to universities to pay for the libraries. But there is a question about whether the government should be directly intervening and providing access, or providing the revenues necessary to do the publishing.

There is value in the content itself, as well as in the value-added services that publishers, disseminators, aggregators, and distributors provide. There is thus a question about how different business models might succeed at supporting both aspects of the process, both content creation and distribution, getting incentives to generate the knowledge in the first place and getting reasonable incentives to provide high-quality publication, dissemination, indexing, and abstracting services. Having a particular business model that may address some of those needs, may not address others.

Finally, the overriding question for this entire symposium—much less this session—is: What impact is the digital publishing world going to have on science itself, on the scientific enterprise? For instance, how are the different business models, the different ways that access is provided and structured, going to affect the quality and productivity of science, collaboration at a distance, access for developing countries, the professional review and career process, peer review, and other aspects of scientific research?

COMMENTS BY PANEL PARTICIPANTS

Brian Crawford, John Wiley & Sons

John Wiley & Sons is a global, independent publisher, established in 1807. Wiley has three major areas of publishing today: scientific, technical, and medical (STM) journals, higher education materials, and professional and trade information. It is a fairly diversified publishing portfolio, with about $1 billion in annual revenues. The publisher has 400 STM journals online on its *InterScience* platform, which was established in 1997. It now contains about 2 million pages of information in the sciences, technology, and medicine.

Finding Wiley's Content Online

Wiley provides open access to its tables of content and abstracts, as a general principle. That is so that its information can be found easily on the Web, either through secondary information services or directly by users who are browsing for such information and coming in directly to the service.

Each journal has its own home page. The reason for that business model is that Wiley publishes on behalf of many learned societies. Wiley wants even those journals that it owns to maintain their own distinct identity online. It also sees the advantage of a rich integration of links to and from related material. HighWire certainly led the way with this, with its toll-free linking capability that they demonstrated with professional societies. Other publishers quickly followed suit. The establishment of

the cross-referencing service was quite visionary. And, of course, all publishers in the biomedical area benefit tremendously by the linkage provided by PubMed. So, the first aspect of the business model is being found on the Web.

Content and Usage Statistics

Wiley has about 250,000 articles across those 400 journals, more than 100 of which are in biomedicine and linked to PubMed. It also publishes laboratory protocols on the Web. It has about 12,000 contributions from major reference works that have been placed online. It also has been moving its books online, about 300 books per year, from a parallel path of print publication.

Wiley's online platform has about 400,000 users who have chosen to register. Registration is optional, and many users do not choose to identify themselves. The service gets about 3.6 million article or chapter views per month, and more than 12 million user sessions per year. The second message about Wiley's business model is that it wants very much to drive usage and to encourage pathways for that use.

A Customer-Pays Model

The next important aspect of Wiley's business model is that it is very much a customer- or reader-pays model of delivering the content. As a result, the publisher places a great emphasis on sales staff worldwide—staff that it did not use in the conventional print environment, where it was marketing via direct promotion to scientists. Libraries customarily took print subscriptions, which did not require sales representatives to meet with them to negotiate electronic licenses. It is an added cost consideration, but important in driving the kind of revenue model necessary to securing the financial success of electronic publishing.

Licensing Considerations

As Wiley embarked upon the development of licenses for its electronic journals, it did so very much in consultation with its major institutional customers. This resulted in flexible sales options. It did not want to make a commitment to one preferred option early in its electronic publishing, so it provided a menu of print and online options.

Wiley now uses basic and enhanced access licenses. Basic means title-by-title access, with some concurrent user restrictions. This is an option that is often suitable for the smaller institution or department. Its enhanced access license does not require that an institution take all Wiley titles that are available electronically. The institution can choose, but it is establishing a license for a larger body of work, with no concurrent user restrictions and with additional benefits such as negotiated price caps.

The publisher determined that individual and society member access privileges had to be instituted in parallel with its institutional sales models, because it also caters to an individual audience, unlike some other publishers who spoke in the first panel session. Its individual member audience is actually holding steady, if not growing, particularly for some of the smaller societies it deals with.

Wiley has been licensing access to journals online since 1999, which has resulted in strong revenue growth. That came as a result of its sales effort worldwide, but also because of the receptivity of the customer base, primarily the institutional library base, to making that switch to electronic resources. This is not an e-only shift though—it is very much a parallel purchase situation. For the most part, institutions are choosing to retain their print subscriptions until there is a reliable, permanent digital archiving solution.

Customer Relationships

The other part of the business model is an emphasis on direct relationships with customers. The business model has shifted away from selling solely through subscription agents as intermediaries. Wiley now often deals directly with its customers. It views its customers as partners, rather than simply as targets. It uses librarians as a channel to end users. Customer training is another mechanism to promote usage, something that all publishers need to do more. Another important aspect is a focus on customer loyalty to sustain business growth. Wiley also provides password access for its society members and individuals, recognizing the roaming nature of its professional customers.

Continuous Enhancements

Finally, Wiley constantly emphasizes investment in new features and enhancements. Some of the most recent ones include content alerts to apprise its audience of what is being published; delivery of content to mobile-edition platforms; and publishing online in advance of print publication.

In order to move beyond a pure subscription model of access, Wiley has developed a feature called Article Select. This is the sale of digital tokens to its licensed customers, enabling them to distribute those tokens to other users within their institutions to access content from journals that they do not subscribe to. The use of those tokens then can be examined to determine what journals might be added to that institution's collection on a subscription basis. This capability has now been expanded into a purer pay-per-view offering.

The Big Deal—Issues in Bundling Licenses

Many not-for-profit organizations are still in the early stages of offering bundled journal licenses to their institutional customers, whereas commercial publishers have undertaken much more aggressive licensing in recent years. This has caused some debate about this so-called big deal.

Wiley's enhanced access licenses now include more than 1,000 institutional or consortia customers. That is not those who sign the license; it is those institutions that are part of such consortia. It constitutes about 70 percent of its e-license business. Wiley's basic access license customers indicated that there are many institutions that are not prepared to make the move to the big deal. More than 3,000 of its customers are still taking the title-by-title electronic access option, which represents 30 percent of its e-license business.

A quick review of the respective license terms is useful to help make the distinction between the business models. The basic access license tends to be to a single site, one geographically contiguous building or campus. There is activated e-access for each subscribed title, and it is for one concurrent user per print subscription if the institution holds a print subscription.

The enhanced access license tends more to be multisite, multinational, and suitable for large consortia, with unlimited concurrent users. There is access to all subscribed current content as well as back content, and privileges such as Article Select, roaming access, and monthly usage reports. Wiley also offers license terms for individuals, what it calls a personal access license, providing individual subscribers with electronic access on a subscription basis; society member access, with members being given access to their subscriptions on terms on which they agree with their societies; and complimentary access to the publisher's chief editors and members of its editorial boards.

Free Access to Developing Countries

Wiley offers free access, free from the time of publication, to developing countries. For its biomedical journals, it does this through the Health InterNetwork Access to Research Initiative (HINARI) program of the World Health Organization (WHO), in order to get content into the hands of investigators in parts of the world where they truly cannot afford such information.

30

Wiley's participation in HINARI, like that of other publishers, is viewed primarily as philanthropy, reaching out to provide information where it is very much needed in order to provide research support and medical care in struggling parts of the world, and not as way of seeding a market. The parts of the world that have been served initially by WHO through HINARI are not seen as being markets in the near or even medium term.

Distinction Between the Paid-Access Model and the Author-Paid, Open-Access Model

Wiley has chosen to maintain the paid-access model described above, and rejected what may be called the author-paid model that is being advocated by proponents of more open access to the literature as an alternative. The reason it has not elected to go the latter route, and why most of the society partners that it publishes with are quite chary of that, is perhaps best articulated by Professor Hal Varian, an economist at the University of California at Berkeley. Professor Varian has pointed out that any economic system tends to favor the entity that pays. Given that the bedrock of scientific communication is the peer-reviewed literature for original work, Wiley's view is that any system that charges the author or a sponsor of the author in order for that author to be published is, in terms of the economic system, going to favor the author's desire to become published. There is nothing wrong with that, but one has to accept the consequences. What has evolved in scientific communication for the most part is a system where the reader pays, or an agent for the reader is paying, because that naturally introduces a selective filter that reassures any entity wanting to scrutinize the validity and the value of the work that it has passed muster.

It does not mean that peer review cannot exist in and support a system where the author is paying. However, one can offer a provocative example of why this is incompatible with the sort of publishing that Wiley does, certainly in biomedicine. That is, one never hears of anyone proposing that a pharmaceutical firm should fund the Food and Drug Administration to evaluate a new drug for approval, because this would be incompatible with the system of checks and balances necessary to prevent bad drugs from getting out on the market. Similarly, we have to be careful that bad science does not get out on the market. That does not mean that work that has been peer reviewed as a result of author support cannot be valid. It does mean, however, that the value of the information would be altered by such a mechanism.

When we talk about value of information, we tend to debate how much the sale value of the information is worth, or about the cost of actually producing that information. In undertaking any study at the National Academies that would look at the value of scientific information, however, one should not assess the value of information on the basis of the cost to produce it. Instead, the value of the information is its value as a tool, as a productivity multiplier in society. One could argue that scientific information, when it is peer reviewed and selectively filtered, has tremendous value to the author, but also to society. That kind of selectively filtered information is that productivity multiplier. Wiley, therefore, has very much supported a customer-pays model for the business because the company believes it ultimately enhances the value of scientific information for those who should value it most.

Joseph Esposito, SRI Consulting[8]

These are heady times for publishers. There is probably more interesting activity in the industry right now than at any time since the paperback revolution following World War II. Unlike the paperback revolution, however, which allowed publishers to grow enormously and to become highly profitable, the current state of affairs is often thought to present more challenges than opportunities for publishers. Indeed, it is not uncommon to hear people say that publishers are a historical artifact, who will be disintermediated by the Internet. Or, we hear that copyright is robbery and that the model for scholarly communications in the future is the open-source software movement, which has given Microsoft a challenger in the form of Linux. When you look at some of the more intriguing projects organized around

---

[8] Joseph Esposito's current affiliation is Portable CEO.

31

intellectual communities, such as the DSpace project that is being developed by Hewlett-Packard at Massachusetts Institute of Technology, even a commercial publisher has to concede that critics of traditional publishing may have a point.

At the same time, commercial publishing employs a great number of people, and it is unlikely that they will all simply go away because a librarian in Nebraska has learned to manage a Web server. It may be useful to consider how publishers see their own activities evolving and what the environment for scholarly communications is likely to look like in 5 or 10 years.

Consolidation of Publishers

From a commercial publisher's point of view, the primary thing to be said about the academic research publishing market is that it is mature, which means that there is only modest growth in revenue, making it hard for new players to enter the market. There are conventional ways to deal with a mature market, and we should expect that publishers will try them all. We should expect, for example, to see the consolidation among publishers continue right up to the point of antitrust action. It is highly unlikely that any scientific discipline will have more than two or three information providers in the years ahead. Some observers believe that there will only be two or three information providers for the entire STM market, covering all areas of study. However, this would result in fewer operating synergies across disciplines.

The form that consolidation will take may change, however. In addition to the already familiar phenomenon of the big companies buying up the smaller ones, we should look for the smaller companies to link together in an attempt to become big companies. We also should expect to see big companies establish distribution arrangements with smaller companies by, for example, putting the content of the smaller companies onto the electronic platforms of the larger companies. This provides critical mass, which can be used to shoulder aside the offerings of organizations with a small number of publications. It is not necessary for Reed Elsevier to buy Sage Communications if they can cut a deal to deliver Sage's publications on the Elsevier platform, but the impact on the marketplace is the same.

Bundling

A mature market will also intensify the downward pressure on prices. This is especially true in academic libraries, where the open-access movement was created in part to put publishers under pricing pressure. Publishers will likely counter this by targeting the market share of other publishers, rather than looking for significant increases in library budgets. In other words, this is essentially a struggle for market share. This is already happening. For example, Reed Elsevier is now offering libraries access to previously unsubscribed journals, not by charging for each journal separately, but by simply insisting in an increase in total expenditures over the prior year. Previous speakers at this symposium referred to this practice as super-sizing, or the big deal. In other contexts, this practice is generally called bundling or tying. Bundling will have the effect of greatly increasing the number of Reed Elsevier publications available through particular libraries, at the expense of having less well positioned publishers lose those customers entirely.

Downstream Value Migration

We also should expect commercial publishers to seek so-called downstream value migration and to target competitors that for various reasons are thought to be vulnerable. Currently, publishers sit in the value chain between authors and wholesalers, and wholesalers have direct relationships with librarians. By moving downstream, more publishers will attempt to disintermediate the wholesalers and reap the wholesalers' marginal revenue. Critical mass is necessary for this, which is another impetus to consolidation. Disintermediation strategies that do not provide significant new value to end users are probably ineffective, but that does not mean publishers will not try it.

Targeting Vulnerable Competitors

What makes one competitor weak and another one strong is a matter of debate, but we are likely to see commercial publishers target the not-for-profits more aggressively in the future. The reason for this is that, rightly or wrongly, the not-for-profits are perceived to be slower to respond and less capable of defending themselves, though they often have substantial goodwill in the marketplace. There are many prospective targets, but one might name just two: the American Chemical Society (ACS) and the Online Computer Library Center (OCLC). The ACS is a target because of its sheer size. The OCLC became a target when it acquired the assets of NetLibrary, making its strategic direction appear to be increasingly like that of a competitor.

Creation of Meta-Content

Meta-content is interesting in that it sidesteps the copyright issue. Content about content can take the form of bibliographies, indexes, and, most important of all, search engines. Publishers are keenly aware of open-access publications and are looking for ways to make money from them. One way to do this is to create search engines for open-access content. Even better is to integrate open-access content with proprietary content for search purposes, as this will tend to bring even users of open content to proprietary services because the breadth of materials included will be much greater. In effect, open-access provides publishers with lower costs for content development even as it makes charging for access fees possible. From an economic point of view, copyright may not be necessary for supporting publishing profits.

The Shift to Web Services

The real economic response to open access, however, is likely to be in the creation of Web services, a dynamic substitute for the publication of fixed content in hard copy. In a Web service, a publisher provides online software that manipulates or processes data that are uploaded to it by a user. The user creates the content and then pays the service provider for the online processing. Think of what it would be like to do your taxes online and you will get the idea. Copyright is irrelevant for models like this, but the economic potential is very great.

Diversification of Customer Base

One final strategy, like many of the other ideas here, is already playing out in the marketplace. It is customer diversification. If the academic channel is mature, publishers will seek new sales channels. The most likely one, because of its size and creditworthiness, is sales to businesses. We should expect to see more publishers trying to modify their offerings for this channel. Since businesses, unlike research universities, are profit-oriented, the kinds of publications they are most likely to purchase speak to an immediate business need. This will tend to shift capital investment away from pure research publications toward applied research and engineering.

Looking out 5 to 10 years from now in the world of scholarly communications, we are likely to see an ample number of open-access publications coexisting with proprietary ones, and we will witness ingenious publishing strategies designed to extract economic gain even in the absence of copyright. The toll collector may have switched positions, but he still has his hand out.

Wendy Lougee, University of Minnesota

This presentation addresses journal costs and value, and the kind of tension that libraries feel as the intermediary between the publishers and the consumers of publications. These tensions may suggest

new models, new ways of looking at revenues. One aspect that has not yet been talked about, but that constitutes a behavioral constraint that should be taken into account, is the longstanding convention of the scholarly journal, going back three centuries. Journals gradually accrued the function of providing a place to document discovery, to disseminate ideas, and also to codify prestige. That kind of three-century convention is not going to be easy to change.

Responses of Libraries to Recent Cost and Marketing Trends

Libraries are the intermediaries between two types of economies. They buy the content in a market economy, and users create and use content in a so-called gift economy. Libraries are often caught in the middle. Libraries have witnessed decades of dramatic journal price increases, with average annual increases of 8 or 9 percent for the past 5 years. It is a very inelastic market, in that as the prices have gone up, the libraries have not necessarily withdrawn or canceled the journal subscriptions. Data over the 15 years ending in 2001 show that journal prices went up 215 percent, yet libraries canceled only about 5.1 percent of their subscriptions. The rates of increase are such that if you look at the Higher Education Price Index, the rate of journal price increases or library materials now outstrips utilities and health care costs. This is certainly cause for concern.

Michael Keller earlier cited last year's report by Morgan Stanley on the STM market. It serves as a primer for investors, shedding light on STM publishing strategies. The report highlighted Reed Elsevier's profit margin of 37 percent for core titles and 22 percent for the medical titles that it recently acquired from Harcourt.

The Morgan Stanley report further discussed three trends worth future attention. First noted was the cyclical slowdown in industry growth due to research library budget cuts, coupled with a downturn in investments. A second trend noted the benefits of scale increasingly accruing to large players. It highlighted, in particular, the kind of bundled portfolio of journals that has come to be known as the big deal, as has already been mentioned. And, finally, it noted that the profit margins will expand for those publishers who have a successful online platform. Interestingly, the report revealed that libraries spend an average of $1.50 on staff and operating costs for every dollar they spend on acquiring content. It speculated that the saving for libraries in moving to electronic-only would be somewhere on the order of 16 percent.

The libraries have experienced a different range of models than publishers, some of which have been discussed at this symposium. These various models frequently come with a requirement that the library sustain its commitment to a certain level of spending, and they give the library very little predictability or control over costs. As the publishers are experiencing these cyclical effects, there is a tendency to stay the course with the expectation that the level of commitment by subscribers will remain stable.

This past year we have seen two trends reflecting an interesting shift. The first is that the publisher strategy of the big deal—the multiyear, all-titles approach—has begun to unravel. We have seen a number of major libraries withdraw from that model, largely fueled by budget reductions and by the recognition that many titles in the bundle simply are not used. The libraries are beginning to push for much more finely tuned licensing models whereby they can select the content that they want—not only at the title level, but in terms of other kind of bundling, much like the tokens that Brian Crawford described.

Moreover, in the big deals, the libraries have encountered pricing that is very hard to defend to their user community. At the University of Minnesota, the library worked with Reed Elsevier to move away from its bundled deal, but found it hard to explain and rationalize the end result: A reduction from 750 subscribed titles to 650 titles, and a move from two formats—print and electronic—to electronic only, resulted in a higher per-title cost.

A second disturbing event this year, which also has been alluded to by several speakers at this symposium, is the demise of some of the other intermediaries, the serial vendors. There was at least one major incident of a serial vendor going bankrupt, resulting in a potential loss of $73 million in library subscription payments. This raises the question of where the functions of procurement and distribution will go. These are functions that have to be sustained somewhere. The likelihood of increased revenues for libraries in the near term (particularly increases that match inflation in journal prices) is low. A recent informal survey conducted within the Association of Research Libraries suggested that nearly half of respondents expected cuts in some areas and the prospects were high for further budget reductions in

34

the coming fiscal year. Library budgets, a major source of revenue for publishers, are obviously stressed. The volatility in the publisher marketplace will probably continue, as will the push from the library community for the more finely grained models that allow them to make some choices.

Implications of Changes in Journal Format and Content

It is also important to look at the changes in format and content, how these changes might affect the existing service models, and the resulting costs for libraries.

There are data indicating that much more intensive use is likely to come with electronic content. There are also recent studies of university users nationwide that have revealed an overwhelming preference for electronic format. Nearly half of all faculty in most disciplines reported they use online materials for the majority of their work. In the biological and physical sciences and mathematics the percentage of information needs met through online resources exceeded 60 percent. Yet interestingly, despite that preference, other studies looking at the perceptions of convenience and ease of use show a dramatic gap in terms of how the library performs in delivering that content. Users are citing evidence of an inability to really control the content, to navigate it well, or to deal with the myriad different platforms and channels that it comes through. Libraries increasingly are challenged to deal with these shortcomings—that is, they are challenged to develop systems to penetrate the so-called deep Web and federate the content in a meaningful way.

An experiment done by the University of California system looked at the differences between print and electronic use by putting the print versions in storage in order to force the use of the electronic versions and to see what would happen. They found that the relatively modest number of requests to retrieve the print versions from storage was by and large because the electronic versions failed to have some attribute or content that was solely in the print versions, or where the fidelity of some of the representation was not as strong. This result was similar to the data collected in the PEAK study at the University of Michigan (in the late 1990s), which provided access to all the Reed Elsevier journals to 12 different institutions and found that 30 percent of the use was for information beyond the text of the articles. This was at a time when many of the publishers were not converting or indexing the whole journal. Some of the major publishers still have been unable to migrate all aspects of a journal (e.g., the "front matter") into the electronic realm.

These issues of journal integrity are important to understand, because they obviously impact how quickly we can migrate to that electronic-only nirvana. They are also important when we think about the kind of unbundling that might be possible and when we look at different parts of a journal and how they are presented.

There is one last point about how content is changing. Perhaps it is a subtle trend, but we are seeing a shift from our concept of publication as product to the notion of publication as process. There are a number of examples where online publications take on an ongoing, dynamic form as users interact with content and comment. For libraries, which are in the business of managing copyrighted, fixed works, that presents a real challenge. It is a challenge, too, to the STM publishing sector in terms of pricing, how to develop models that support something that is not fixed or well bounded.

There is a marvelous book by anthropologist Sharon Traweek about going to live among the physicists. The analysis (called *Beamtimes and Lifetimes*[9]) sheds light on how the physics community migrated into the e-print form. It describes how the early, informal sharing of information (now represented in e-prints) functions to document what is going on, as opposed to the formal publication that occurs months later, which is seen as a ledger to give credits and debits to fellow scientists.

The Changing Role and Influence of Libraries

As the divide between content and publication begins to occur, what is the role of the libraries? How do they manage and sustain formal publications and more informal processes? What influence do they have in affecting change with publishers, or with scholars and researchers for that matter?

---

[9] Traweek, Sharon. 1992. *Beamtimes and Lifetimes: The World of High Energy Physicists*. Harvard University Press, Cambridge, MA.

The Morgan Stanley report suggests the potential for reduced operating costs for libraries—no journal check-in, no binding, no claiming lost issues. However, the counter-argument might be that the necessary infrastructure to support the investment in that content, to federate it appropriately, to ensure its longevity, and to archive it is going to be an increased expenditure on the library side. What we are likely to see is a shift of any savings to supporting the additional infrastructure. It underscores how critical it is to have community attention to this issue of infrastructure, to interoperability, and the kinds of protocols that will allow that federation to happen.

What influence might libraries have? It is rare that collective action is robust and sustained. A recent exception occurred a few years ago when *Nature* proposed to embargo content for institutional subscribers. Economist Ted Bergstrom has suggested an analogy with the economic construct of coordination games. He has a wonderful parable where he talks about the annual meeting of anarchists. They are used to going to the same hotel, but all of a sudden the hotel decides to raise its price, and they do not know how to coordinate action to change the venue for their meeting. Like the anarchists, librarians and scholars may be stuck in the context where it is impossible to agree on or organize collective action. And, when they do attempt coordinated action, they are accused of collusion.

Libraries do have a role, however, in seeding and developing alternative, competitive publishing models. Libraries have a fundamental understanding about content, its use, and users. There are a number of examples, of which HighWire Press is certainly one. There also is an increasing incidence of institutions, such as Cornell and the University of Michigan, starting incubator services, production services to help small publishers move to electronic publishing. These projects represent a move by the libraries away from their traditional role of providing access to information toward facilitating production of information, and it may help them reconceive the relative position they may have in the STM information sector.

To summarize, libraries play an interesting role in the middle in a time of shifting costs and resource constraints. The near-term focus for libraries is likely to fix on achieving a more finely tuned approach to acquisitions, while at the same time developing the necessary infrastructure for the future.

Patrick Brown, Stanford University

The Public Library of Science (PLoS) is a new scientific publisher based on an open-access business model, sometimes also called a midwife model. Bruce Alberts and James Duderstadt have already pointed out that scientific publication has traditionally been viewed as a public good by the scientists in universities, the sponsors of research, and the citizens whose tax dollars pay for much of the research that gets published. The goal of scientific publication from the perspective of these stakeholders has always been to make the results of scientific research as widely available and useable as possible. None of the important constituencies are benefited by limiting access or the freedom to use the published scientific information. This is the operating premise.

Reconceptualizing the STM Publishing Business Model on the Internet

Before the Internet, the hard reality was that there was no choice but to charge the readers and users of scientific publications, because the most efficient way of distributing this information was by printing it on paper and delivering it in boats and trucks. Every potential user thus represented an incremental expense for the publisher. So, of course, any business model that did not take that into account was doomed to lose money in proportion to the popularity and success of its publications. That system, however, had the unfortunate consequence of limiting distribution to individuals and institutions that were willing and able to pay, and this was accepted as a necessary evil. With the advent of the Internet, that model is no longer a necessary evil.

The traditional business model also had another indirect consequence that was more subtle, but equally unfortunate. It was based on selling research articles. This gave rise to the mindset that the content of scientific journals was valuable property that the publishers were somehow entitled to own, control, and sell for a profit. This is a really big psychological barrier that we now have to overcome.

With the worldwide spread of the Internet, something fundamental has changed about the economics of scientific publication, not just the technical means of distribution. There is an absolutely

categorical change in the economics that is wonderful, because it makes Jefferson's ideal of the infinite, free dissemination of scientific ideas and discoveries a realistic possibility. What had been an impossible ideal in the pre-Internet era—to make the published information an open public resource—is now possible because the cost to the publisher no longer scales with the number of copies produced or with the number of potential readers of a publication. That means that we are not restricted to a business model that charges per access or per copy. In fact, it means that a business model that restricts the distribution and use of the published work is really working against the interest of science and society. It is unnecessary, anachronistic, and inefficient. This business model now stands in the way of the ideal that we have been working for all along, and that is now potentially available to us. If we do not need to charge for access, then we should not charge for it.

Open Access Defined

The PLoS will use an open-access business model. An open-access publication is one that meets the following two conditions. The first is that the copyright holder (either the author or the publisher, if the copyright has been transferred to the publisher) grants to the public a free, irrevocable, perpetual right of access to the work, and a license to copy, distribute, perform, and display the work and to make and distribute derivative works in any medium for any purpose.

The second condition is that not only do they grant that license, but they actually make it possible for people to access the work. They do that by making a complete version of the article and all supplemental materials available in some suitable standard electronic format, deposited immediately upon publication in at least one internationally recognized, independent online repository that is committed to open access. One example of such an open-access archive is PubMed Central, maintained by the National Library of Medicine. The PLoS plans to make everything that it publishes immediately and freely available online, through its own Web site, PubMed Central, and any other distributor or library server that wishes to house and distribute its journals.

Under this definition, "open access" does not mean just some free peeks or downloads from the publisher's Web site. It means that everything PLoS publishes can be downloaded in its entirety, manipulated, copied, redistributed, integrated with other data, and even resold by commercial publishers. The point is that the content is not the publisher's property. The publisher in this model plays the role of a midwife, an expert at delivery that gets paid for the service but does not expect to keep the baby.

Advantages of the Open-Access Model

The advantages of this kind of true open access are already very familiar to everyone in the life sciences. They are certainly not theoretical. In the life sciences there are two longstanding, amazingly successful experiments in open-access publication, GenBank and the Protein Structure Database. The success of the genome project, which is generally considered to be one of the great successes of recent times scientifically, and perhaps even its existence, is in no small part due to the fact that the world's entire library of published DNA sequences has been an open-access public resource for the past 20 years. If the sequences could be obtained only in the way that regular published work can be obtained, one article at a time under conditions set by the publisher, there would be no genome project. The incredible value of genome sequences would be enormously diminished.

More significant is the fact that open access is available for every new sequence, and it can be compared to every other sequence that has ever been published. The fact that the entire body of sequences can be downloaded, manipulated by anyone, and used as a raw material for a creative work has meant that thousands of individual investigators on their own initiative have taken up the challenge of developing new data-mining tools. It is such tools and the new databases that incorporate sequences, enriched by linking them to other information, that have really made the genome project the big success that it is. With open-access publication, there is absolutely no reason to think that we would not see at least as great a flowering of new, investigator-initiated research using that wealth of published material as a resource for creative work, and adding to its value.

37

Open Access Supported by the Author–Pays Business Model

Unlike the subscription-based model discussed by the previous speakers, to pay for the cost of publication the PLoS plans to charge the cost to authors and their sponsors. From the standpoint of business logic, this is by far the simplest and most natural model. It is natural, because the cost of online publication is scaled with the number of articles, not with the number of readers. It also makes perfect sense from the standpoint of institutions that pay for research. Their mission is to promote the production and dissemination of useful knowledge. From that perspective, publication is inseparable from the research that they fund.

The PLoS initially plans to charge about $1,500 for a published article, with no charge to authors who cannot afford to pay. That figure is based both on a ground-up analysis of the publisher's cost at steady state and also on a survey of publishing costs from a number of publishers who were kind enough to open their books to PLoS.

Research sponsors should welcome this model, because for a fraction of 1 percent of the cost of the research itself, the results can be made truly available to all, not just to the fortunate few who are at the lead research institutions like Stanford that can afford to pay for site licenses. It can include the students and faculty of community colleges, high schools, and smaller institutions; scientists in poorer countries; and patients—and there are plenty of them—who want, on their own initiative, to read about the latest progress in research on their condition and are unable to pay exorbitant fees to read all the articles. This model therefore would greatly increase the benefits that are realized from the public research investment.

In the short term, of course, there will be an incremental expense, but in the long term, once the scientific community has made the transition to open-access publication, there will actually be savings to the major research sponsors because they are ultimately the ones who pay for most of the costs of the subscriptions in libraries and even for individual subscriptions. The Howard Hughes Medical Institute (HHMI) is one of the largest funders of scientific research in the life sciences, and it has already endorsed this model. The HHMI has agreed to provide budget supplements to its investigators, specifically to cover author charges for open-access publications.

Those of you in the publishing business should take this as a carrot. The money is available only for publishing through an open-access mechanism, as defined above. It is deliberately intended as a gesture of how important HHMI considers open access and as an incentive for the researchers to choose that system. At least some other major funders of research in the life sciences are taking similar steps, and this should be encouraged.

The PLoS is also going to produce printed editions of its journals that it will sell to institutions or individual subscribers at a price that is intended just to recover the cost of printing and distribution, everything downstream of producing the published digital document. This is estimated to cost on the order of $160 a year, but that will depend on many things, including the size and the distribution of the journal, as well as other factors. If there is a market for the printed journal, which there very well may be, then the print journal operation is not going to be selling the content, it will just be selling a physical form of the content that some people may find convenient.

There will not be any cross-subsidies between the open-access online publication and the basically break-even print publication operation. That makes perfect sense, because now that these two versions are being decoupled, the system is much more flexible. The PLoS does not care how well the print publication business does. The market will decide whether it is something people value.

Concluding Observations

In closing, it is useful to re-emphasize a point that was made by the first two speakers at this symposium—that scientific publication is seen as a public good. This ideal of scientific publication being managed as a public good, as an open public resource, is now within our reach. If we, as scientific publishers, can work together, and work with the universities and the institutional sponsors of research, we can make the transition from the anachronistic, restricted-access business model that we have now, to a system that is economically sustainable, much better for science and society, and that provides free and open access to everyone.

DISCUSSION OF ISSUES

Relative Merits of the User-Pays and Author-Pays Models

Pat Brown began with a comment for Brian Crawford. He thought that Mr. Crawford's argument in favor of restricted access, namely, that it had an important role in maintaining the quality of publications and that an open-access publication model would somehow undermine quality standards in scientific publications, was specious. There are two grounds for saying that. First, all the studies that he knows of, and there have been a number of them that have looked at the relationship between the price and measures of quality of scientific journals—citations, their assessment by peers in the field, and so forth—have found, overall, a dramatic negative correlation between price and quality. The notion that users paying for journals somehow upholds high-quality standards is not supported by the data.

Second, every scientist, at least in the life sciences, knows that most of the journals that are regarded as third-tier journals of last resort, but which publish 99 percent of the articles, have no author charges but very high subscription costs. The premier journals, however, typically wind up charging authors in color charges and page charges something on the order of $1,000 or more. So, all the evidence that Pat Brown knows of actually tends to argue against the point Mr. Crawford made.

Brian Crawford responded that Professor Brown may have extrapolated incorrectly from what he had said. He did not say there was a correlation between the price and the quality or value of the information. He was making the distinction between two different economic models—one where the customer or the user is paying, versus one where the author is paying. Dr. Crawford's point was that there is an inherent bias in a system where the author pays, which ultimately would devalue that information in terms of its overall utility as a productivity multiplier.

There are two reasons for this. The first is that a less selective filter would be imposed. And second, going back to Hal Varian's argument, is that an economic system where the author pays is naturally going to favor the author. That means that any entity wanting to make decisions about that work needs to impose yet another filter at some cost, in order to determine how that work stacks up against others.

Pat Brown added that in his view, that second filter exists already, and there is no added cost for it. The factor that serves to maintain the quality of the work that authors submit for publication is not that they do not think they can ever get a paper published. One can always find a mediocre journal that will accept just about anything. Rather, it is that the authors know that sooner or later their peers are going to read what they write, and their reputation depends on it being good. That is ultimately what determines their career advancement, their status in the field, and so forth. So, that filter is already in place. It is not at the level of whether or not a paper can find its way into print.

Brian Crawford said he thought that most scientific publishers would say that selectivity is probably not exercised all that often by authors, because they seek to get their work out. Professor Brown did ask another question regarding the subvention of publishing costs by the payment of page charges or subvention for color reproduction charges. This is a good point, and an example of where some organizations have found an effective balance. They have done this even in a user-pays model, implementing certain charges that are passed on to the author, where the needs of the author are seen as unique and something that the author would want to pay in order to benefit from a service. Most organizations, however, do not make that a criterion for a decision of acceptance or rejection; it is merely a matter of presentation of the work.

Steven Berry, of the University of Chicago, said that Hal Varian's point that publishers would favor the author by charging the author is absolutely right, and that this is exactly what the scientific community would like to do. The public good that is produced by scientific research is very special. In contrast to other public goods, the value of scientific public goods increases with use, and it is specifically in the author's interest. It is very much a part of the motivation of the scientific paper author to have that paper as widespread as possible, more and more used, thereby enhancing the value to the community, as well as the interests of the author. Consequently, Brian Crawford's quote of Hal Varian is very much an argument for placing the charges to the author, rather than to the user, who does not have a particular vested interest in increasing the use of each written piece of work.

Brian Crawford clarified his remarks further. It is absolutely correct that one needs to optimize two

39

variables. One variable is the dissemination, and the other is the filtering. His point was that the author-pays model moves the filter boundary, whereas publishers also have an interest in disseminating the work as broadly as possible. It is a matter of getting the balance right between those two considerations.

Professor Berry added that despite Mr. Crawford's logic, the facts already indicate, as Pat Brown observed, that the quality of articles in the journals is not correlated with the users' charges; if anything, it is inversely correlated.

Donald King noted that one of the things that he has found in his collaborative research is that it costs authors on average about $7,000 to produce an article. Roughly one-third of the articles are funded by the federal agencies, about a third of them are funded exclusively within the universities, and the rest by industry.

There are two problems in this context. One is with the argument that Steven Berry was making. How will the funders of these authors be convinced that they need to pay an additional $1,500 above the $6,000 or $7,000 that they are already paying the authors to prepare those articles?

It is not an easy argument to make for all of those different sources of funds. For the government funding sources, such support can be written into the grants, although that has its own problems. About 75 percent of the readers of STM articles are outside the universities and the granting organizations, so they may not want to pay for page charges, especially if federal funding for research grants is reduced. Nevertheless, some good arguments can be made here. Where you are going to have even more difficulties, however, is with the people who are writing those articles, who are paid by the universities or industry. You will need to have different kinds of arguments for those three different constituencies. How are you going to get all of them to do it?

Pat Brown said he thought that the interests are similar for all three of those constituencies, namely, whatever motivated them in the first place to support the research and to encourage the authors to publish it. The motivation to pay an increment of less than 1 percent of the total research cost to make it much more valuable to the people who are supposed to be served by it is presumably the same for all three sectors that sponsor research.

Donald King continued that it may be necessary to go further in trying to understand the funding priorities and budget profiles in each sector. For example, one could look at what the universities are already purchasing through their subscriptions and get some notion of what the transfers of funds are. Once you have those data, you might be able to have even stronger arguments for the author-pays model.

Ed Barnas, journals manager for Cambridge University Press, said he has been involved in both scientific society and commercial journal publishing for about 30 years. One thing about the discussion of business models that has been ignored thus far is the question of the mission of the organization that is doing the publishing. That does have a significant impact on the business model selected.

For-profit, commercial publishers will seek to disseminate information in journals that are very attractive to the library market, the institutional market that provides them with the financial return on those journals. Not-for-profit, nonsociety publishers will seek to disseminate information also, but in a nonloss situation so that they can continue to stay in business and serve whatever their own mission is. Society publishers, however, have a significant investment and interest in serving their members, and in that regard, some of the comments about author support, page charges, and color charges may be seen in a different light.

When Mr. Barnas was working for the society publisher, or working with societies as a contract publisher, various societies would make use of the page charges as a way of supporting their member subscriptions, keeping the member rates low, so that they could provide more benefits to their members. A commercial publisher or a university press cannot really charge page charges, unless it is a journal they are publishing on behalf of a society because that is not viewed as appropriate. It would be seen as gouging. Color page charges might be acceptable, because they are a direct cost, but author page support for articles in non-society publications cannot be a viable part of current business models for non-society publishers.

Brian Crawford commented that he considered the points made by Mr. Barnas to be quite valid. One should not look at the customer-pays model, as opposed to the open-access model that has been proposed as an author-supported model, in absolute terms. Indeed, many professional societies have had a hybrid model, where they have benefited from subscriptions at the same time that they have used author page charges and other subventions to help support their publishing programs and to keep the costs to the customer down. That kind of hybrid model has been determined over time by market forces.

40

One could ask, Why are we debating now what the right business model is? Many of the business models that we have serve us well, and have been adopted over time.

The Effect of Different Publishing Business Models on the Long-Term Preservation of Digital Journals

Wendy Lougee raised the question of how publishers see the incorporation of the long-term preservation requirements in the context of the user-pays economic model, given the unknown future costs associated with that.

Brian Crawford noted that the focus of publishers in electronic publishing is mainly on the actual production of the work and its dissemination on the Web. They have not addressed the important issues that deal with long-term preservation, including the integrity of the information and its migration to new platforms. Reed Elsevier has begun a very sensible project with the Royal Library in the Netherlands, where they are looking at how to address those kinds of long-term preservation problems. They will be costly, and the costs will need to be shared between publishers and libraries, ideally with support from governments and other institutions.

Jeffrey MacKie-Mason asked Pat Brown to address that same question from the perspective of the PLoS. Professor Brown's definition of open access included the immediate deposition of the publication with some organization committed to long-term access. However, there is no guarantee the organization will follow through on this commitment. One of the advantages of print publication in is that with enough copies produced, very long-term preservation is more or less guaranteed, which is not always the case with electronic publications. The PLoS will make print copies, so does it plan to deposit print copies in archival institutions as well?

Pat Brown responded that PLoS has a number of plans for providing archival stability for its information. What is important, however, is not only that somewhere there exists a permanently preserved copy of the information, but that it is permanently openly available to everyone. Of course, one would be hard put to find a more trusted and trustworthy archival repository than the National Library of Medicine, but the fact that the information is also going to be freely and widely distributed, and will exist in many institutional servers, provides another measure of assurance.

Wendy Lougee added that her question was focused more on the natural process involved in digital preservation—the durability of format, whether it will migrate, whether the publisher is doing all the right things to guarantee success—and then how to incorporate those requirements and their costs into the publishing business model. Will it be ensured through an escrow fund or folded into the base cost, and how will those costs be accounted for realistically when they are unknown in advance?

Pat Brown noted that it would be a terrible mistake for institutional repositories to agree to take on the job of archiving information without requiring the publishers to grant unrestricted open-access rights. To have them take on the burden of preserving information that the publishers still treat as their property, in his view is really an abuse.

Martin Blume, editor-in-chief of the American Physical Society, addressed the archiving issue, because his society has put all of its content going back to 1893 online. It is linked and searchable, and PDFs are also available. Current content is added as it becomes available. There will be costs in changing the format of this in the future. The society also recognizes the concerns of libraries and of the entire scientific community that this historical record might be lost. What would happen if the American Physical Society were to go under?

There is a full mirror site of the entire archive that is already accessible and tried at Cornell University, which is where *The Physical Review* originated back in 1893. The full content of that, from 1893 to 2001, is 378 linear feet of shelf space. The journals are also deposited at the Library of Congress. If the society is terminated, it has an agreement with the library that these holdings will be put in the public domain, freely available to anyone. This is the society's primary effort for ensuring continued availability of its archived publications. It should be noted, however, that unlike the American Physical Society, Cornell is not an eleemosynary institution. The society, therefore, has provided the university with servers for keeping its digital publications available and also pays a fee.

Issues in Transitioning to the Open-Access Model

Martin Blume also noted that the American Physical Society has an open-access model for one of its journals that it started five years ago—*Physical Review Special Topics, Accelerators and Beams*. It is a small journal, but it is available completely without access barriers. It is not, in fact, free because it does cost something to publish it. The costs are recovered by subscriptions through the sponsorship of 10 large particle accelerator laboratories around the world. It is also cosponsored by the European Physical Society's Interdisciplinary Group on Accelerators. At this point it is only a limited experiment because the American Physical Society has other expenses that it must recover, and it has to see whether these things work or not. The society cannot afford to bet the whole ballgame on the open-access approach because it still needs to recover its costs.

In a published debate in *Nature* on this topic, Dr. Blume did propose that the society would like to put everything online without access barriers. It could do this right now if every organization that now subscribes to its journals would make those subscriptions sponsorships, providing enough to recover its costs. Then it could be opened to everyone else. The risk would be that institutions might be tempted to decide that since all journals are available, they no longer have to pay anymore. The libraries would love that, but then the publisher would go under. If the society were to attempt this, and did not recover enough sponsorship in the future, it could have a "pledge week." Every time a reader tried to download an article, there would first be a solicitation for contributions.

Doris Peter, from the *Medical Letter*, a not-for-profit tertiary publisher that publishes reviews of drugs approved by the Food and Drug Administration (FDA), expressed concern about pharmaceutical influence in publishing, and industry influence on publishing generally. She wanted to know if PLoS is going to accept sponsorships from industry.

She said that there appears to be some support for industry to be more involved in supporting STM journals. From the *Medical Letter*'s point of view, it purposely maintains its independence because it could not credibly publish independent reviews of drugs if it were to accept money from anyone that is associated with the pharmaceutical industry. In addition, Brian Crawford earlier cited a hypothetical situation in which the FDA might accept money from the pharmaceutical industry. In fact, the pharmaceutical industry does pay the FDA to review new drugs.

Pat Brown responded that he completely agreed with Ms. Peter's point of view and that it would be a bad idea to accept such sponsorships. It is not something that the PLoS is contemplating doing.

Nick Cozzarelli, editor of the *Proceedings of the National Academy of Sciences* (PNAS), said that the idea of open access is a great one and he hopes that PNAS someday will be switching to that business model. The problem is that PNAS is a not-for-profit journal, with no endowment. The risks of going immediately to an open-access model would be too great.

His question to Pat Brown was whether he saw any paths or way stations between full open access and the current subscription-based model. Dr. Cozzarelli suggested that one possible approach might be for a publisher to give the author an option for a surcharge. The author could pay extra to have immediate open access to his or her article, whereas all the other articles of authors who did not pay the surcharge would be free in PNAS after six months. Are there any other ideas of some sort of possible intermediate path?

Pat Brown responded that he did not object to the surcharge idea in principle. The amount of the surcharge, however, must not be so high as to be a disincentive. It is not a crazy model, although he does not think it is the best way to publish. He realized the difficulty of making that transition. The PLoS had to get a large grant from the Moore Foundation to buffer the financial risk for its experiment. Without the grant, the PLoS could not have done it.

In addition, Martin Blume's suggestion in the *Nature* debate was not a bad idea as a transitional approach. In fact, it was something Professor Brown would strongly urge the PNAS to consider, that is, to ask the institutional subscribers that now provide most of the revenues, mostly academic institutions that probably accept the philosophy that journals provide a public service, to continue to pay their subscription fees at the current rate for some interval of years to be specified, during which time the PNAS would make the transition to open access. With such a multiyear commitment of support, PNAS would have a stable revenue source that is not put at risk by making that transition. It can try to make the transition, at the end of which time it can determine whether it looks like it is going to be a self-sustaining model.

42

Professor Brown thought that would be worth doing. He would hope that the current subscribing institutions would not take the low road and try to undermine the process by saving themselves a little money. They could see that it is in their own best interest in the long run to encourage open access. He asked whether Dr. Blume's society had actually proposed that to institutional subscribers, at least as a transitional approach.

Martin Blume responded that the American Physical Society is having some discussions with other U.S. organizations to start this. Of course, the situation is different in every country of the world. He added that this is not unlike other consortium agreements that are made, whereby well-off institutions offer to pay a little more, and then open access can be provided for all institutions. Unfortunately, he has found that even the big institutions have asked, "What's in it for us?" and have refused to go along with this approach. They are unwilling to help the smaller organizations. So, the society has made efforts along this line, but so far to no avail. Such an arrangement is not unlike what Ohio Link does, but they came up with some extra money that enabled the society to open its journals to the institutions in the entire state of Ohio. That is the kind of catalyst that is needed, a bit of extra money.

Pat Brown noted that it might be best to view this approach as transitional, because ultimately the sensible thing would be for the research sponsors to cover the publication costs as an essential part of their mission of promoting and disseminating research. In the short term it is necessary to catalyze the process. He asked if the libraries would be open to trying an openness approach as a temporary experiment.

Mr. Blume answered that the problem with the temporary approach is, What happens if it does not work? It is very hard to get people to resubscribe. Once a library has given up the subscription and used the money somewhere else, resubscribing becomes a new acquisition.

Pat Brown added that the sponsoring institutions could be provided an incentive to maintain support. For example, if Stanford provides $5,000, or whatever the amount may be, then everyone affiliated with Stanford would have open-access publishing rights in the journal. Therefore, it is a competitive advantage for Stanford to offer this. If you treat it as kind of a credit pool that could be drawn on by authors from the subscribing institution, then even in the open-access model, they are getting some special benefits beyond what the nonpaying institutions are getting. It becomes an added incentive.

Constraints of the Print Model on the Electronic Model

Bruce McHenry, of Discussion Systems, raised the issue that everybody is still working in the paradigm of the old paper model of publishing, where there is a lot of prepublication work, because a big print run is needed in order to economize on the costs of distribution. The questions about archival policy, editorial policy, and open access all change completely if one moves to a model of continuous improvement of the materials, or continuous publication, where all peers have an opportunity to adjust the prominence of newly developed pieces. In that model, the world changes completely; this issue is discussed in more detail in the last two sessions of the symposium.

Pat Brown noted that a prerequisite for the kind of innovation that Dr. McHenry has raised is that the materials that are being continually re-evaluated are an open public resource. It is hard to see how this kind of approach would work, as a practical matter, if you still have a system in which every publisher's body of work is in a separate, restricted repository.

Bruce McHenry commented further that there are certain domains outside academia where that is simply not viable, because people need to be able to earn a living for doing intellectual work. In that case, they cannot make it free and open right away. However, after some period of time, as the prices go down, foundations, government, and corporations may decide to purchase the materials and make them accessible within their group or publicly.

Advertising Revenues in Electronic Publishing

Professor MacKie-Mason raised a question posed by a Webcast participant who works for a large society publisher. Nearly 50 percent of this publisher's total costs are attributable to editorial, peer-review, and production processes—items other than printing, paper, and distribution. Advertising associated with paper journals underwrites more than 50 percent of the total cost of operation. If there

43

were less advertising in the electronic-only model, then although it may save on some costs, it actually would represent a major lost source of revenue. Would electronic-only publishing be a viable business model if you were not to charge even more than publishers currently do for electronic access that is coupled with a print version?

Brian Crawford responded that for journals with broad member or individual circulation, most advertisers still see print as their means of reaching that audience. They are not yet ready to move away from print advertising to online-only in scholarly publications, although this may not be the case with other consumer-oriented publications. It would be very risky to go to an online-only strategy for STM publications, if the publisher currently relies on such advertising revenue.

Sponsor-Supported Open-Access Model

Steven Berry raised the issue of the sponsor-supported publication model, in contrast to the author or the author's institution supporting the publication. He believes that many researchers like the former model because it is the sponsor of the research who shares with the author the interest in having the product disseminated as widely as possible.

When the National Institutes of Health (NIH) supports research, it is because it is going to be published and made into a public good. Consequently, to fulfill the goals of supporting the research at all, the sponsor assumes a responsibility to see that the material gets published. For this reason, the HHMI model that Pat Brown described is a very good one, and many scientists who publish research would like to see that kind of approach propagate. Is there a way to persuade the less well funded agencies, particularly the National Science Foundation and Department of Energy, to take on the responsibility to pay for publication? This would be in contrast to the present model, which puts the author payment responsibility into a grant in which the author has discretion whether to spend that money for publication or to spend it for support of another graduate student or some other research cost.

Pat Brown answered that despite his best efforts, he is not sure if there is a way to persuade them. It is an open question. However, one can make a strong argument that the publication costs amount to less than 1 percent of the research costs in biomedical research, (although he does not know the percentage in other disciplines). If that is true, if you had to take a 1 percent hit on other aspects of the budget to do it, it is a plausible argument that the return on that investment would be extremely high, as opposed to the 1 percent cut from other areas, because all the grantees and all the research that an agency is funding would be providing much freer access to a more extensive body of information. Again, the purpose of funding of research is not just to serve the immediate community of the grantees, it is the wider scientific community and the general public that should be much better served by the information.

At the same time, when this issue has been raised with respect to NIH funding, many NIH grantees have objected to such a proposal if it were to come at the expense of a 1 percent cut in research funding. They argue that there is not enough research funding to go around now; so it obviously is a controversial proposal.

Author Selectivity in the Open-Access Model

Ted Campion, senior deputy editor at the *New England Journal of Medicine,* commented that open access makes eminent sense, especially for anything of archival value, with GenBank being the ultimate example. Scientific information should become freely open to all. At the same time, the value and the future of any publication are going to be determined mainly by what authors want to do, where authors want to publish. The NEJM is open to all authors; 4,000 scientific articles are submitted to the journal each year, with no charge for submissions and no page charges if the paper is accepted. However, the NEJM does reject many submissions. The role of a biomedical or scientific journal is to be critical and selective. That is in part why authors want to publish there. If authors or their sponsors have to pay, is this selectivity going to be compromised? How is the PLoS going to exercise the functions of peer review and of being selective? Is that part of its model?

Pat Brown agreed that the journals most attractive to authors tend to be the ones in which they have the least chance of having their papers accepted. The PLoS certainly has factored that into its development strategy. It intends to be very selective from the beginning—not selective just on the basis

44

of whether the article is good enough to be published somewhere else, but selective on the basis of publishing papers likely to be of interest to a very wide audience as well, precisely because the PLoS considers this important in terms of developing the journal identity and as a magnet for submissions.

If the point was that adding author charges would be a disincentive to authors to submit to the NEJM, that would be highly unlikely. Professor Brown has looked into author charges quite a bit and has paid plenty of them to a lot of journals that are very selective. It is not at all uncommon for such charges to be more than $2,000 per paper. There is a prototypical archival journal in the life sciences, for example, that does not charge for submissions, but it does charge for color. It is $750 for the first color figure and then $500 for the second. About 20 percent of all the papers that are published in that journal (in the several issues that Professor Brown has seen) had, on average, two color figures. The authors were willing to pay extra money to add color to their articles published in a completely archival journal. In the vast majority of those cases, the colors were primarily to make it look better and not because color was essential for the scientific content. The point is that an author will not balk at paying extra money after putting a great deal of work into a scientific paper for all the world to see.

If we go to a model where the institutions are covering page charges as an essential part of research, it will make it even less of a disincentive to authors. In Professor Brown's view, the *New England Journal of Medicine* has nothing to worry about in terms of going to author charges, and such a change would be better for the community it is supposed to be serving.

Bob Simoni, associate editor of the *Journal of Biological Chemistry* (JBC), which is a publication of the American Society for Biochemistry and Molecular Biology, commented that the JBC charges authors page charges. On average, an article in the journal costs about $1,000, which includes both page and graphics charges. This has allowed the journal to keep its subscription price quite low.

The journal also considers itself to be an open-access journal, though it does not comply fully with the current PLoS definition. Two years ago, prompted by all of the discussion that came up with PLoS, the JBC decided to find some way to make all of its articles free online as soon as they are published. The journal started something that is called PIPS, or papers in press. On the day a manuscript is accepted, the paper is published on the Web in PDF. It is free to everyone, and it stays free forever. Anyone with Internet access can read freely every single JBC article that has been published. It has been enormously popular with the journal's readers and authors. The fact that it is published on the day of acceptance reduces the time to publication by eight weeks on average, which is the time usually taken for copy editing and processing; so it has served everyone well.

45

5. LEGAL ISSUES IN PRODUCTION, DISSEMINATION,

AND USE

INTRODUCTORY COMMENTS

Jane Ginsburg, Columbia Law School

The discussion in this session is divided into three areas: the basics of copyright and ownership of rights; licensing by authors to publishers and from publishers to users; and some of the economic and non-economic rewards to authors from scientific, technical, and medical (STM) publishing.

Before addressing the issues surrounding the grant of copyright from authors to publishers, it is important to review the "upstream" question of whether academic and research authors are initial copyright holders, or whether the institutions that employ them in fact hold the copyrights. The initial position matters for a variety of reasons, which need to be better understood. With regard to rights: What rights does copyright confer on authors (including the right to make available for free), and what rules govern the transfers of rights?

In terms of licensing, what rights can authors realistically give to or withhold from publishers? What are some alternative forms of managing rights, such as broad license to publishers while retaining copyright, and what are the pros and cons? What are the major issues between the publishers and users of STM publications and of various models of access? Institutions generally gain access to electronic information via license. What kinds of access do these licenses offer? How could access be improved? What are the advantages and disadvantages of licensed access models compared with hard copy distribution? What role might the science funding agencies or other organizations play in resolving conflicts in relationships between authors and publishers, or publishers and users?

Finally, what are the current and potential economic and non-economic rewards to STM authors from publication, in both traditional and alternative venues? How else, apart from monetary compensation, can authors share in the value of their work? The Social Science Research Network (SSRN), an online publisher of abstracts and (when authors can provide it) full text for working papers or forthcoming published articles, is used as an example to explore these and other related issues in more detail.

Copyright Basics: Ownership and Rights

To begin this discussion, it is useful to examine three questions about copyright and scientific, technical, and medical publications. First, what rights does copyright establish, in what kinds of works? Second, who owns copyright in scholarly works? That is, does the academic author own the copyright, or does the university that employs her own the copyright? And third, assuming that the academic does own the copyright, what rights does she end up having to give up in order to get published?

Basic Rights Conferred by Copyright Law

With respect to copyright basics, first of all, copyright protects works of authorship, and that is a very broad notion. For example, it includes scientific works of all kinds, including computer software, whatever the medium of expression. Whether it is on paper, on film, or in ones and zeros, if it is a work of authorship, it is protected by the copyright law. It is important to understand, however, that while copyright protects the work, it does not protect the information as such within the work. Copyright does

46

not protect ideas or facts, such as genome sequences. It only protects the article that describes or expresses the results and analysis of that information, but not the information itself.

With respect to the exclusive rights that copyright confers on the copyright owner, those include the right to reproduce the work in any medium, including digital; the right to make what the law calls a derivative work, that is, the right to adapt the work, including updating the work or making further works based on the first work; the right to distribute the work in copies, including digital copies; and a right that is becoming of increasing importance on the Internet, the right to publicly display the work. When you see the work on your computer screen that you are viewing from a Web site, that is a public display of the work, and the rights of the copyright owner may be implicated thereby.

It may come as a surprise to a lot of people that copyright does not grant a right of attribution in the United States. There is nothing in our copyright law, unlike the copyright laws of many other countries, that gives the author the right to be recognized as the author, that gives the author name credit. This is a great failing of U.S. copyright law. To the extent that authors enjoy this right, they enjoy it either by contract with those who publish their work, or possibly as the result of other laws, but not the copyright law. While the research and academic communities respect a right of attribution in practice, such a general right does not exist under American law.

The exclusive rights under copyright also are not absolute. There is the well-known fair-use doctrine, which derogates from the exclusive rights. Copyright is limited in time as well. The term of copyright in the United States is now the life of the author plus 70 years. That is a very long time, and it is the same term in the European Union.

Ownership of Copyright

Copyright ownership vests in the author of the work. That makes a lot of sense until one finds out that according to the copyright statute, the employer of a work for hire is considered to be the statutory author. What is a work for hire? For present purposes, a work for hire is one prepared by an employee within the scope of his or her employment. That sounds bad to most academics, since it might follow that everything they write belongs to the university, at least if you apply the statute's terms literally.

In fact, there has been a long tradition predating the 1976 Copyright Act under which professors own the copyrights in what they write—their articles, books, lectures, and teaching materials. There is no indication that Congress ever meant to change that tradition in the 1976 Act. It just wrote an all-purpose provision on employee authorship transferring authorial status, as well as all copyright rights, to the employer. The question has not been settled under the current Copyright Act, whether the rule is what the statute says or whether the long-standing practice remains valid. Two appellate judges in the 7th Circuit, which sits in Chicago, have suggested that surely Congress did not intend to disturb this long-standing practice and understanding. At the same time, one has to acknowledge that the two judges who said that are Judges Posner and Easterbrook, both of whom have been very prolific professors, as well as judges, so their statements may not have been devoid of self-interest.

As a practical matter, the question of professorial or academic copyright ownership did not even come up, apart from these speculations about the text of the 1976 Act, until relatively recently, probably because most academic copyrights, as opposed to patents, were not worth much. Although universities have had patent ownership policies for a long time, most universities, until recently, have not had any kind of policy on copyright ownership.

That has changed in the past few years, in part because of the rising importance of software. Software is protected by copyright, and it is worth something. So, some universities thought that perhaps they should lay claim to software—not only software written by staff, but also software written by professors. Universities traditionally had distinguished between works by administrators or staff on the one hand, which were works for hire, and works by professors, which were not works for hire. Nevertheless, some universities began to lay claim to software produced by their professors.

Another thing that changed was the continuing escalation of journal subscription prices. Some universities thought that perhaps if they owned the copyrights, rather than the actual creators owning the copyrights, they could bargain better with the publishers. That was perhaps more speculation than actual practice.

Finally, the most recently flurry of copyright ownership policies was precipitated by distance education on the Internet. Before the dot-com bust, a lot of universities began to get stars in their eyes

47

about the commercial potential of online distance education, and saw this as a potential profit center. They therefore wanted to make sure that the content that was being distributed via online digital distance education would belong to the universities, and not to the teachers. At least, that was their first impulse. The allure of online distance education has diminished since those policies were first being adopted, however, and many universities appear to have backed off from asserting such claims.

Table 5-1 presents a selection of copyright ownership policies that were publicly available from major public and private research universities. These are divided first into universities whose default position is that faculty own the copyrights in their works, followed by universities whose default position is that the university owns the copyright in faculty works, most notably on grounds of work for hire. For those researchers who are contemplating moving from one university to another, there could be valuable bargaining information here about faculty ownership or not of their works. What difference does this actually make?

TABLE 5-1 Default University Ownership. Source: Jane Ginsburg

	Default Ownership Position: University Ownership	Conditions For Faculty Ownership	Written Agreements	Exceptions to the Conditions for Faculty Ownership	Creator Acknowledgment/Right of Withdrawal (When University Owns Copyright)	Faculty Role in Subsequent Disposition of University-owned Work	Transfer or License Back to Creator	Provision for Revenue Sharing
University of Arizona	University owns work of employees, including faculty.	"Employee-excluded" works (traditional publications in academia, artistic works, academic software) owned by creator.	Creator responsible for executing documentation to assign ownership and secure protection of property owned by the university.	Board retains paid-up, nonexclusive license of "employee-excluded" works for education, research and public service. Board owns IP created by visiting faculty, subject to exceptions on a case by case basis consistent with the policy.			Release if in best interest of university; discretionary license to creator. Release subject to nonexclusive license to U for educational, research, and public service work; minimal royalty; cannot use university resources to improve the invention.	University will pay creator a share of net income, (gross revenue minus administrative fee and unreimbursed costs). Percent paid to creator determined by "IP Official" in accordance with university policy. Revenue sharing continues beyond employment.
University of Chicago	University owns "the intellectual property the faculty create at the University or with the substantial aid of its facilities or its financial support."	Individual faculty enjoy the revenue "until it is substantial," at which point, university and faculty member share. Revenues will generally be too minimal for it to be "appropriate" that the university assert ownership rights.						"Individual faculty members are entitled to share in the revenues from the intellectual property they have a hand in creating." Share will vary, depending on faculty member's contribution and costs incurred by university.
Emory University	University owns all work of "Emory Personnel" related to normal duties (including areas of research/scholarly expertise) or using Emory support.	Emory does not assert rights to "academic or scholarly Copyrightable Works," including "works for hire."	Emory personnel must agree in writing to be bound by policy; personnel receiving substantial support must sign specific IP Rights Agreement Form.	If "related to Emory Personnel's normal duties...field of research or scholarly expertise;" specifically assigned; or used substantial support.			Can petition for release or assignment of ownership. University retains nonexclusive, royalty-free, perpetual license for research, clinical, and educational purposes.	Contributors receive 100% of gross up to $25,000; thereafter revenue is split with department, school/center, and university generally. Contributor will receive his/her personal share after departure or death.

49

Institution							
University of Minnesota	University is sole owner of "all intellectual property created through use of University resources or facilities..."	"Regular academic work product," as specifically defined, is an explicit exception and thus owned by the creator.	Written agreements required to effect exception to the regular academic work product rule of ownership. University has an IP Policy Acknowledgement form.	Regular academic work product assigned in writing to the university; specifically ordered/commissioned work designated in writing as such.		No express provision in policy for transfer or license, but policy stipulates that nothing should affect right of creator to publish, except for delay as needed for university to secure protection.	University will share revenue derived from "academic research or scholarly study," and "other intellectual property" as deemed "appropriate" after consultation with creator's supervisor. No mention of share after creator leaves university.
Ohio State University	University asserts ownership rights in "Products of University Research."	"University Research" excludes works of "artistry, academic instruction, or traditional scholarship."	Necessary prior to beginning work on a "specific assignment."	Otherwise excluded works created as part of a sponsored program or resulting from a "specific assignment."	If University exercises rights, university will pursue protection and licensing "with the assistance of the…creator."	If university does not arrange, within a reasonable time, commercial development, transfer to third party, or dedication of rights to the public, creator may request transfer or waiver. Subject to nontransferable, royalty-free license for university.	Creator shares in royalties derived from any "Product of University Research." No explicit mention of rights after employment ends, but policy provides for university share if creator cannot be located within five years, which implies ongoing rights.
University of Texas	Board may assert ownership in intellectual property of all types.	Board "shall not assert ownership" of works by faculty, unless commissioned or hired specifically to create the work.	General Counsel "shall prepare and distribute... model agreements" as appropriate for implementation of policy. Creators encouraged to sign an acknowledgment clarifying ownership of works they create.		Creator should have "major role" in publication decision, but discretion rests with university regarding development and commercialization.	University may choose not to assert interest; may elect to release after asserting interest, subject to possible recovery of expenses, reserved income rights, or other limitations.	No share for commissioned work or work created by employee specifically hired to or required to produce it; for other university-owned work, university "recaptures" costs, then splits remainder with creator.
University of Virginia	Acknowledgment that work-for-hire gives the university ownership of works by employees.	University "cedes" ownership and thus will not assert ownership interest in "copyrightable scholarly and academic works" created by faculty with use of generally available resources.	Required when university cedes ownership of a work in which, under the policy, it would otherwise retain an interest (not required for scholarly and academic works).	Use of university resources *and* work generates royalty payments or is of commercial value that can be realized by University marketing. University retains a nonexclusive, royalty-free right for noncommercial use of works in which it "cedes" ownership.		University may cede ownership of works in which it otherwise, under the policy, would have retained an interest. Retains a nonexclusive, royalty-free right for noncommercial use.	Income is distributed according to guidelines developed by the Vice President for Research and Public Service.

50

| **Yale University** | Begins with work-for-hire premise that works created by faculty in the course of their teaching/research are university property. | Disclaims ownership of works by faculty, staff, postdocs, and students. | Assigned tasks and special circumstances of unusual commitment by university to project (discretionary). | University will share "net income" at its sole discretion, but generally on the same terms as patent policy. University deducts costs for securing and defending copyright, and for licensing or otherwise using the work. |

51

About two-thirds of the universities that were reviewed start out with the position that faculty own the works that they create, subject to a number of fairly typical default-shifting conditions. Not every university has all of them, but most of them show up in one form or another.

First of all, if the university has invested substantial resources in the creation of the work over and above resources commonly made available to faculty, then the university will assert ownership in that work.

Second, many universities have a category of institutional works, which may border on administrative, or often refer to courses that are uniform throughout the university, such as the Columbia Contemporary Civilization Course, which has been taught and added onto by generations of faculty and is considered really more of a university work product than the work of any individual faculty member or faculty members. Those are also the objects of university assertion of copyright.

Third, many universities provide that where the research was sponsored by an outside entity, if a condition of the outside entity is that the university own the copyright, then the university will own the copyright.

Fourth, a number of schools will assert copyright ownership if the work qualifies as "technology" or software. This can be troublesome, because most of these categories are not well defined between "traditional" academic work, which belongs to professors, and "technology" which belongs to the university. There is some potential tension in the understanding of what those categories mean.

When the default position is that the university owns, there is often a provision that the university will forbear from asserting copyright in traditional faculty works. That leaves open the question: What do we mean by a traditional sort of faculty work? If it is not clear what that technology category means, and if the university's position is default university ownership, it may be that the faculty members are on the short end of that stick.

Regardless of who starts out owning the copyright, most policies have some kind of limitation on the exercise of copyright, whether by the faculty member or by the university. The most typical constraint on the faculty in the exercise of copyright is that the faculty member will be asked to give the university a royalty-free license to make nonprofit university educational use of the faculty member's work product. Often that license applies even after the faculty member has left the university.

On the side of university ownership, if the university owns, whether by default or because one of the conditions for transferring ownership back to the university applies, the most typical constraint on the university's exercise of copyright is to allow faculty noncommercial use rights. Some universities will not commercialize the work without obtaining the reasonable consent of the faculty member in advance. Furthermore, many universities provide that with respect to course materials, even if the copyright is owned by the university, if the faculty member leaves and goes to teach somewhere else, he or she can take those materials to the next place.

However, only one university, Columbia, provides that even when the university owns the copyright in the work, it should respond favorably to the creator's request to the university to make the work freely available. In other words, the university in effect renounces its exercise of copyright control over the work that it owns at the behest of the creator of the work. This provision extends not only to faculty creators, but also to staff creators. The provision was added to the Columbia policy largely at the request of the academic computing staff, who do not necessarily hold faculty appointments, but who felt very strongly that what they do should be made freely available for other people to build upon. The university acceded to their request. Other universities may do this as well, but it is not explicitly included in their policies

Rights That a Faculty Member Gives Up To Be Published

Assuming that the faculty member does own the copyright, what does he or she have to give up in order to get published? Table 5-2 presents the copyright policies for authors from a cross-section of publishers, both the not-for-profit professional societies and for-profit publishers. It is rather interesting that there is not much difference in the terms between the societies and the commercial publishers.

TABLE 5-2 Default Faculty Ownership. Source: Jane Ginsburg

	Default Ownership Position: Faculty Ownership	Conditions Changing Default (Excepting sponsored projects)	Other Limitations on Faculty Copyright	Limitations on University Ownership	Written Agreements	Creator Acknowledgment / Right of Withdrawal (When University Owns Copyright)	Faculty Role in Subsequent Disposition of University-owned Work	Transfer or License Back to Creator (if Conditions for University Ownership Exist)	Provision for Revenue Sharing
University of California	Scholarly/aesthetic work resulting from "Independent Academic Effort" by a "Designated Academic Appointee" owned by creator.	Contracted facilities work; commissioned work; sponsored work, and project (unless otherwise provided); institutional work.			Required for Contracted Facilities Work, commissioned work, sponsored work, and project with © ownership other than as provided in the policy.			Discretionary, subject to third-party agreements and best interest of university. Work may not be further developed with university resources; university is granted free-of-cost, non-exclusive, license for educational/research purposes.	Royalty or income from third-party assignment or license of university-owned copyrights may be shared with creator based on "originator's contribution, University's costs," agreements with sponsors, or other applicable agreements.
Columbia University	Recognizes faculty ownership of traditional works of authorship created by faculty.	Substantial use of or support by university resources; created or commissioned for use by university; work-for-hire (including administrative duties). University may assert copyright in: institutional works; course content / courseware (possible exception); use of university name; software.	In licensing work to third parties, creator shall seek to reserve university royalty-free right to use portion of work within university for teaching, research, noncommercial use.	If university holds rights, faculty members can use works for noncommercial purposes.	Policy constitutes binding understanding; formal agreements as necessary to implement policy.	Acknowledgment of creators who make "a substantial creative contribution" (including creators of work-for-hire).	Creator of a work owned by the University can request the work be made publicly available. Creator has consultation regarding commercialization.	Transfer possible upon university determination that ownership has no commercial value. University retains irrevocable, royalty-free license for noncommercial purposes	Revenue shared for works made with substantial use of university resources; sponsored project; courseware. Share on a case by case basis for institutional works and works-for-hire. Share not altered by termination of university affiliation or death.
Duke University	Personal ownership of intellectual property rights in works of the intellect by their individual creators.	Computer programs; databases; extraordinary allowance, grant, or subvention; certain collaborative works.	Members of university community enjoy permanent, nonexclusive, royalty-free license to make customary and traditional use of course content.			Moral rights of individual creator "respected to the extent practicable in every case contemplated by this Policy."		Discretionary grant "of licenses or royalties or both…on just and reasonable terms."	Royalties granted at university's discretion.

53

Harvard University	Except as qualified, "a member of the University is entitled to ownership of copyright and royalties or other income derived from works."	Work created by non-teaching staff (work-for-hire); use of rare university holdings (potentially); substantial university assistance; use of university name or insignia; extensive use of voice or image of students or staff.	If work falls into optional exception categories, and creator continues to be owner (excluding work-for-hire), university may share in royalties and enjoys royalty-free or reduced-royalty license for not-for-profit educational activities.	Explicit agreement necessary when there is substantial university involvement (even if agreement is that creator retains copyright).		Share does not apply to work-for-hire. For other university-owned works, creator's share continues after individual leaves university. Creators deriving "substantial income" encouraged to continue tradition of making a gift to university.
University of Illinois	Creators retain ownership to "traditional academic copyrightable works" unless subject to defined exceptions.	Creation pursuant to external agreement, as a specific requirement of employment or assigned duty; commissioned work; patentable work.	If resources beyond those "customarily provided" are used, creator copyright subject to perpetual, nonexclusive, royalty-free license for use in internal teaching and research.	Acceptance of policy deemed condition of employment. Commissioned work must be prepared under written agreement.	Introductory principles advocate consultation for creators in the publication, development, modification, or commercialization of university-owned work.	Assignment in case of university abandonment; license upon demonstration by creator of capability to commercialize. Assignment must grant university perpetual, nonexclusive, royalty-free license for internal teaching/research; may be subject to revenue sharing. License subject to conflict of interest review.
Johns Hopkins University	"Copyright to, and royalty from, literary or scholarly works" produced as part of "usual teaching, service, and research" belongs to creator.	University has "right to obtain title to IP developed as a result of support either directly from or channeled through the University." An appointment itself is not "University support."			Work usually excepted from university ownership but work that is a "specified deliverable" from a university-funded or -sponsored project is owned by university.	Creator shares in "net revenues." Personal share survives termination of affiliation with university and death. Creator does not share in revenue from sponsored research agreement or grant, nor is there a share in equity comprised of real or personal property. Other revenue is shared with creator, subject to deductions for "out of pocket payments or obligations" plus a reasonable reserve.

54

Institution						
Massachusetts Institute of Technology	Author ownership subject to carved out exceptions.	Work-for-hire by operation of copyright law; works created with significant use of MIT funds or facilities. (Policy does not state whether faculty work is generally considered work-for-hire.)		Must sign agreement if participating in sponsored research, MIT-funded research, or "significant use" work.	Subject to university discretion, author may request that work be placed in the public domain.	Discretionary license back or "stand aside"; creator must demonstrate ability to commercialize. Royalties to university applicable. Transfer may be terminated if effective dissemination does not occur within three years.
						Provision for share after deduction of overhead and direct out-of-pocket costs.
University of Michigan	"All faculty own and control instructional materials and scholarly works, created at their own initiative with usual University resources."	Work created with "unusual university resources" or "as a specific requirement of employment."	Nonexclusive use and distribution rights for noncommercial purposes, continuing after faculty member's departure from university. First refusal to make new versions of university-owned instructional materials (consultation for those who left university).	Policy states written agreements "may be necessary" to modify rights or provide greater specificity.	Acknowledgment and withdrawal.	Right of first refusal on new versions; consultation on reuse and revision of instructional materials for creators who have left university. Departed creator may use work for teaching, research, and noncommercial purposes.
University of North Carolina, Chapel Hill	Ownership dependent on category; "traditional" or "non-directed" works are owned by the creator. ("Directed works" are works that are "specifically funded or created at the direction of the university," whether or not exceptional resources invested.)	If exceptional resources are used, university owns copyright. If university is involved in commercializing traditional/nondirected work, assignment required.	As a condition of employment, faculty grant university nonexclusive, nontransferable, royalty-free license, unless such will impede scholarly publication.	In case of commercialization by university of creator-owned work, an Assignment Agreement shall specify responsibilities and proceed shares.		Release possible in case of work using exceptional resources, directed work, or sponsored work. Nonexclusive, nontransferable, royalty-free license for university's own education/research; income should be shared with university.
						Policy states that royalties for copyrighted works are usually a matter between author and publisher; acknowledges that "different treatment" may be necessary for university-owned works but does not specify terms of such treatment. Likely will accord with patent policy for net revenue sharing.

55

University								
Pennsylvania State University	Ownership of "literary, scholarly, and artistic works" rests with creator. (Such works do not include "survey instruments, instructional materials including videotapes, and computer software created on U time.")	Works generated within scope of creator's employment (literary, scholarly, and artistic works specifically excluded) or commissioned by university.			All faculty must sign an Intellectual Property Agreement.	Policy Manual has a provision on coauthorship of scholarly reports; applies regardless of copyright ownership.	Provided only for patents.	Policy provides for sharing after recovery of costs related to prosecution, maintenance, or infringement litigation costs incurred by university.
University of Pennsylvania	Affirms academic custom of creator ownership in works resulting from research, teaching, and writing.	Substantial use of university resources; work-for-hire, defined in policy as prepared at direction of supervisor, pursuant to job description, or administrative duty.	In transferring interest to others, author should attempt to secure for university royalty-free right for traditional, customary, or reasonable academic use. University retains non-exclusive, royalty-free license for software and courseware.		University may request assignment in instances of work-for-hire; written agreement necessary for works making "substantial use" of university resources; will attempt to "secure acknowledgement" from authors in sponsored project.		Faculty may petition university to waive ownership of work created with substantial use of university resources.	Revenue from sponsored research and work involving substantial use of university resources follows that of patent policy, but with different share percent; creator share continues following employment at university.
Princeton University	No claim to work produced by faculty through "normal teaching and research efforts."	"Substantial resources... designated for the development of intellectual property" *and* by which faculty member may derive outside income.	Software subject to special disclosure requirements with respect to commercial potential.	Even with the use of substantial resources, university only considers it has "some equity" and "part" of the income should reimburse university for use of resources.				Policy provides for sharing of "Net Income."
Stanford University	"Copyright shall remain with the creator"; specific disclaimer with respect to "pedagogical, scholarly, or artistic works."	Institutional works (direct allocation of university funds, created at direction of university for specific purpose, or works without a discrete author). Significant use of university resources requires assignment of title to the university.			All faculty must sign University Patent and Copyright Agreement.		Discretionary reconveyance provided there is no real or potential conflict of interest for the creator, or conflict with university policy. Reconveyance may not "limit appropriate University uses of the materials."	Policy states that royalty will be distributed in accordance with patent licensing policy, which provides for distribution after deductions for overhead and direct expenses.

56

Institution							
Vanderbilt University	Rights in "Literary and Artistic Works" are granted "to faculty, staff, and students who are the authors."	Nonscholarly literary and artistic works created with use of university "funds or facilities, or [which] capitalize" on university affiliation. "Technology" (computer programs, designs) created with use of university facilities, or without use of university facilities but within creator's scope of employment.	"Technology" that would otherwise be owned by university but assigned to outside entity by creator or created pursuant to independent research, consistent with university policy.		Anticipated that creator will participate in licensing, including consultation prior. Affiliated licensing may be given first negotiation.	If university does not patent or transfer "technology," creator can request assignment. Must be consistent with existing governmental rights.	Net income shared with inventor. Purpose of policy is to aid inventors and creators for small discoveries, while large inventions aid inventor's school proportionately more. If inventor leaves university, inventor's *school share* increases by 10%.
University of Washington	University faculty, staff, and students retain all rights in copyrightable materials they create.	Commissioned works (developed during regular duties, except scholarly works); "sponsored" works making extensive use of university service centers.	When necessary for transfer of university-owned or university-sponsored work. Required for commissioned works, for works with uncertain ownership interests prior to beginning, or to vary default status. Required to dedicate author share of revenue of university-sponsored works to school/department.	Can request materials be withdrawn from internal use if revisions are not feasible and provided author is still employed by the university.	Internal: approval of author required unless waived in written agreement or implied; can request revisions prior to internal use. External: subject to written agreement between university and author.		No share in commissioned works; revenue shared in the case of "University sponsored" works, subject to recovery of costs.
University of Wisconsin (policy primarily pertains to instructional materials with brief disclaimer of scholarly copyright; no other policy located).	University does not assert ownership in "traditional teaching, research and scholarly activities," though performed as an employee.	To determine ownership rights in instructional materials developed with minimal university support; also if university will own copyright because of substantial support, assigned duty or work-for-hire.	Agreements under the policy should address internal use rights and sharing, which is encouraged.	Rights re: revision and withdrawal with respect to instructional materials only. Includes consultation, option to assume responsibility for revision; consultation on "other" uses. Author role in external distribution.		Revenue from "external distribution" shared; split based on pre- and post- amortization schedules. Author's share continues regardless of termination from university.	

57

Typically, the publishers demand a total assignment of copyright from the authors. This is pandemic. For example, the copyright assignment form of the National Academy of Sciences (NAS) states:

> I, the undersigned, hereby irrevocably grant and assign exclusively to the National Academy of Sciences my rights, including copyright interests, in materials or work that I will contribute to the preparation of the report identified below, and in the report, under all laws, treaties, and conventions throughout the world, in all forms, languages and media, now or hereafter known or developed without limitation.

Many of the agreements do, nonetheless, provide for some kind of privilege back to the authors, notably for their own nonprofit, university use in further works or in teaching. Where there is considerable divergence is with respect to electronic rights. Almost all of these journals publish both in print and in electronic form. Depending on the contract, authors will be permitted to post an abstract of the article with a link to the publisher site, or they may be allowed to post a preprint of the article on their own Web page, or on some other preprint site, but they may not post the text as edited by the journal. Another variation permits posting of the article, but only on restricted access sites. So, with respect to what authors can do with their own articles, it is far from open access. Another fairly typical provision, also from the NAS contract, paragraph four, states:

> In making this assignment, I understand that it is the National Academies' policy to grant permission to me to use materials prepared for this report without fee upon application to the National Academy Press.

That means that the author cannot use his or her own material without written permission from the National Academies.

One might think these policies of the publishers would create a lot of public opposition by the authors, but they have not. So, why not? Or, at least, why not yet? The simple and completely unlegal answer to that is that professors in practice largely ignore these contractual limitations, and they boldly post full text regardless of what their contract says.

A highly unscientific survey of Professor Ginsburg's colleagues in the sciences at Columbia University suggests that this reflects what may be called the Nathan Detroit model of e-publication: "so sue me, sue me, what can you do to me?" The authors are basically betting that the STM publishers will not actually enforce that contract against them.

A number of the preceding speakers have envisioned an evolution away from the production and distribution of copies on the part of publishers and toward value-added services such as peer review, or what Wendy Lougee referred to as the codification of prestige, as well as toward a variety of secondary indexing and searching services. If it is true that the commercial value will no longer attach to the actual content of the work, which may well remain on the academic's Web site or on a preprint server, then Joe Esposito may be correct that copyright will not matter anymore to publishers.

Copyright nonetheless does matter to authors. One thing that was not mentioned regarding the university copyright default rules is that there is a significant psychological factor in being the owner of the copyright in your work. The universities that take the work-made-for-hire position—to the extent that their faculty know about it—risk antagonizing those faculty members. There is something very deep-seated, even if one is disposed favorably to open access, to still being considered to be the author and copyright owner of your work. Whether or not it will matter to publishers in the future, it will most likely continue to matter to authors.

LICENSING

Ann Okerson, Yale University

Licenses are not entirely new, but are a more recent method for distributing STM journals and other digital information. Licensing can and does work, and this presentation will discuss why. At the same time, there are many people in academia and at libraries who have expressed negative views about the licensing of journals.

Copyright is a long-established regime that most participants in STM journal publishing view favorably. This is also because copyright law is arrived at through a broad national dialogue, and it takes time for that dialogue and for the resulting law to be completed. Copyright is federal national law, and it tends to be very high in concept, and very low in specifics. It is also hard to change, which is probably a good thing.

Licenses or contracts are quite different. They are private, negotiable agreements. They are specific and very tailored, and that specificity can be very reassuring. If you are an elementary school teacher and you like to teach the kids in your class, and you want to know what you can do, a license can help delineate that in very useful ways.

Licenses can restrict or expand rights that would be granted by copyright law. In that sense, a license is a neutral document controlled by the parties to the agreement. Licenses used to be regarded as entirely controlled by publishers, at least in the library community, but this is no longer true in all cases. Libraries, working with publishers in the license environment, have in fact made an electronic market possible.

Why Do Publishers Need Licenses If They Have Copyright?

People often ask why is it that we cannot just use copyright for the distribution of electronic information resources. Why is it that we are now using licensing as well? The most obvious reason is that there is no artifact to work with in the electronic environment. There is nothing to hold, and there is no object to own. This is a very big factor in library licensing.

The electronic distribution environment is very new, and current copyright law does not fully address copyright owners' concerns. Some of those concerns are based on the ease and rapidity of distribution of copyrighted works on digital networks, but there are others.

A number of vendors and aggregators have added value to public-domain databases originally produced by government entities and have found customers willing to pay for such enhanced information products. Copyright may not adequately protect such products.

Other vendors, aggregators, and publishers have digitized a lot of out-of-print works whose copyright has expired and have added a great deal of functionality and value to those works. Again, because the works they have digitized are in the public domain, copyright law would not really help them in protecting their distributed material.

Another reason there is value in licensing is that copyright has been in a state of flux. Even though it may take a long time to change the law, in almost every session of Congress in recent years there has been some copyright or database protection legislation or some other proposed statute concerning rights in digital information. Licenses provide certainty to the parties regarding the specific terms and conditions of the use of the material.

Copyright law also cannot address the details of complicated electronic information arrangements anymore than it could in print. Licenses can clarify and make less ambiguous numerous matters where practices are unclear. They consequently encourage dialogue, which may be the most valuable part of the licensing process. Licensing engages the two parties in a conversation about what kind of expectations they have of each other. Over the past few years in the Internet information environment, this has contributed to a great improvement in understanding between librarians and publishers.

Finally, licenses generally are able to cross national boundaries successfully. They have enabled progress in the marketplace for electronic information without waiting for copyright or database legislation

to be revised or implemented. In the United States, licenses are governed by state law, not national law. They can be harmonized through the National Conference of Commissioners on Uniform State Laws (NCCUSL), although attempts to do so in this specific area of the law have failed thus far.

Major Licensing Issues for Libraries

Not all licenses have been fair or negotiable for libraries. In fact, 10 or 15 years ago, the licenses offered to libraries were unacceptable. Some problem areas have been resolved, but others persist. The most difficult areas to reach agreement on have been the terms for interlibrary loans and guaranteeing perpetual access and archiving.

The interlibrary loan is a relatively long established practice whereby a user in an institution that does not have a book or article can request it from another library. Articles are generally photocopied and sent to the requester by fax or by mail. Now that articles are available online, one might expect that interlibrary loans should be able to take place with the supplying library transmitting the article electronically to the requesting user. In fact, very few publisher licenses permit this type of transmission. Most require that articles be printed from the e-version and then forwarded, or they do not permit any interlibrary loan from the electronic version at all.

When the interlibrary loan from the e-version is only permitted after it is printed, the main consequence is that delivery is more cumbersome than it otherwise would be. Where the interlibrary loan for the e-version is not permitted at all, then a serious degradation of access ensues in the electronic environment, especially if libraries move to electronic-only subscriptions, which, contrary to what was said by some previous speakers, they are. In the latter instance, libraries cannot fulfill interlibrary loan requests. The requesting reader may be unable to afford a pay-per-view solution at the price set by the publishers, or, more likely, the pay-per-view option just does not exist.

The concern about archiving and perpetual access in licenses is somewhat different from this, but it is nonetheless a great concern. Research libraries have indeed attempted to hold onto their print subscriptions for their archival and preservation value and to adopt electronic subscriptions to journals for reasons of service and functionality. However, the economic climate is such that many libraries, for lack of resources to support both print and electronic subscriptions, are beginning to drop the print. This is happening even in the biggest research libraries.

Libraries have always assumed that the material they pay for will last indefinitely, and users will be able to have long-term access to it. The movement to electronic-only suggests two requirements. One is that repositories of electronic journals must be robust, even though currently they are less than optimal. The other is that libraries licensing journals on behalf of their users will want continued access to those e-resources even after their paid access to a given journal stops for some reason.

The majority of contracts for electronic journals now provide language for continued access, often in perpetuity, but not all do. Furthermore, the contracts are very weak because of the inability of most publishers to follow through technologically or in a business sense on the promises that are made about the archiving provisions. These major flaws in archiving provisions will be resolved only upon significant discussions and investments by all the stakeholders about archiving responsibilities, costs, and technologies. Meanwhile, as libraries are now canceling paper subscriptions, the official scientific record is at some considerable risk.

Licenses That Serve Author Interests Better

Scientists clearly want their articles to be widely available. Many scientists today might like to distribute their own articles on their Web sites, their university sites, their lab sites, and e-print sites, in addition to the formal peer-reviewed journals. Some publishers permit this, even though copyright has been transferred to them, but others do not. In most cases, publishers ask scientists to transfer their copyright, and scientists are accustomed to make such transfers and sign those papers.

How can authors get out of this quandary? It is not really so difficult or impossible, and there are two ways of doing this. A good way is to sign a copyright transfer that permits the author to distribute his or her own work. The advantage of this to the author is that it keeps them out of downstream copyright management problems.

60

The copyright transfer form of the American Physical Society (APS) provides one good model.[10] The most important points are that the author has the nonexclusive right to give permission to third parties to re-publish the print version without obtaining permission from the APS. The author also has the right to use all or part of the article on his or her Web home page, or for lecture or classroom purposes. The author also has the right to post and update the article on e-print servers, if certain reasonable conditions are met.

An even better way, perhaps, is to maintain copyright, while licensing the publisher to deploy the work in ways that the publisher needs for effective publication and dissemination. In this case, the author will have continued responsibility for managing his or her rights, but will also have broad flexibility as owner of the work.[11]

ECONOMIC AND NON-ECONOMIC REWARDS TO AUTHORS:
THE SOCIAL SCIENCE RESEARCH NETWORK EXAMPLE

Michael Jensen, Harvard Business School

The Social Science Research Network (SSRN) was not founded to make profits, but as a way to change or to make efficient the distribution of work in the social sciences. The SSRN posts abstracts to full-text, nonrefereed working papers on its Web site, as well as full-text, refereed articles and journals of established publishers who want to have access to the community that has been created around the Web site. The network provides both Web-based access to this work and e-mail journals, which carry abstracts. The announcement service is very important and highly valued. It is one of the few services for which the SSRN charges, and not very much at that.

The copyright policy is very simple. The SSRN does not take copyrights. All it requests is a nonexclusive license to post the electronic versions of the abstracts or working papers on its Web site, and the author can take them down at any time.

The network was originally founded in 1993 to publish just abstracts in financial economics. It began to carry full-text working papers in 1997. It outgrew financial economics very quickly and then launched different related ventures, including the Economics Research Network, the Accounting Research Network, the Legal Scholarship Network, and, most recently, the Management Research Network.[12] It will probably continue to expand.

The SSRN publishes more than 400 e-mail subject-oriented abstracting journals to narrow down the torrent of material to something that is manageable for people. The e-mails show up on each subscriber's desktop weekly, containing three to five abstracts, with links to the full-text paper, full author contact information, and the ability to communicate with each other. The SSRN charges about $10 per year for the e-mail abstracting journals.

Another way the SSRN gets revenues is as a subcontractor to many universities and research institutes that contract out the job of distributing their research. Currently there are 92 institutionally sponsored research paper series, which have the full-text papers. Each of those institutions has its own e-mail abstracting journal, along with the whole associated infrastructure.

The network has an increasing number of established publishers that are giving it their work so that they are part of the database, part of the e-mail abstracting service. They sell the full-text downloads, and share the revenues. The SSRN only charges for the downloads if the partner does.

The size of the network community currently is about 40,000 subscribers; 31,000 authors; 1,800 partners in publishing (both formal publishers and universities or research institutes); 716 academics worldwide who act as advisory board members on these journals and networks; 137 editors and network directors, most of whom are unpaid; and 34 paid administrative and professional staff.

There are about 200 journal publishers with roughly 1,000 journals, for which SSRN has permission to carry abstracts and full-text documents. The nonprofit publishers, which include the

---

[10] The American Physical Society's copyright transfer form can be found on their Web site at http://forms.aps.org/author/copytrnsfr.pdf/.
[11] Several examples of such licenses can be found at www.library.yale.edu/~/license/authorss-licenses.shtml.
[12] See the Social Science Research Network home page at: www.ssrn.com.

universities, research institutes, and similar organizations number about 400, with about 400 journals and research paper series. Partners include MIT Press, Yale, UCLA, Stanford, Skadden Arps, National Bureau of Economic Research, Goldman Sachs, and Blackwell. There are approximately 53,900 papers in their database, all with abstracts, and full-text documents for about 33,000 of them.

The SSRN gets about 3 million hits and about 120,000 unique users per month on its Web site. The average length of a session on its site is 4.5 minutes, which is well above the norm in the Internet context, which typically averages about a half a minute. There are 190,000 downloads of full-text documents per month now. It has had about 4.3 million downloads to date and about 12.6 million abstract views.

The SSRN has created some measures of popularity. Although it does not referee anything, it will put material on the site if it is part of the worldwide scientific discourse in the field for which it is intended. It creates for each field and each journal, "top ten" lists, based on use. The top paper has had 30,257 downloads. This results in an unbelievable amount of attention for papers, at least compared with what you would normally expect to get through a journal. Another top 10 list is for the most popular papers published in one of the network's abstracting journals in the past 60 days.

Every author in the system has an author page. Wherever the author's name shows up on the SSRN Web site it is hyperlinked to a stable URL that takes the user to the full contact information for the author. The author page provides a list of all of each author's papers on the SSRN site, which are hyperlinked and available for downloading. The author page gives the total number of downloads from all of the author's documents, the articles are ranked, and the downloads for each one are given.

In addition to charging small amounts for the abstracting journals, the SSRN charges institutions that want to stop distributing the work themselves and transfer that function to the network. It also carries a professional announcement and job series, for which it charges nominal amounts. In cases where people choose to sell their work through its site, prices range anywhere from $5 per download for many full-text journal articles, to $20 a download for a couple of the publishers, and they share those revenues with us. Although the SSRN is not yet self-financing, is expected that it will eventually cover its cost.

DISCUSSION OF ISSUES

The Value of SSRN to Authors

Jane Ginsburg began the discussion by giving the SSRN a warm endorsement as a SSRN author. She noted that in the field of legal scholarship, it has made a tremendous impact through those abstracting journals, which arrive regularly. You can select what fields you want to know about in recent filings with SSRN, and then click on the full text. More often than not you can get the full text, and not just the abstract.

Then there is the author gratification element of SSRN, which is not generally available elsewhere. The brilliance of SSRN in developing those top 10 lists is that it has devised an infinite number of carefully gauged categories so that almost every SSRN author at one point in his or her career is a "top 10" SSRN author. The author then gets a nice e-mail from Michael Jensen conveying congratulations that the article has been selected as one of the top 10 in some category, even if it is an extraordinarily narrowly defined category. It is very gratifying. Authors can look at their Web page and see how many downloads there were of their article in any given week.

That is one of the reasons that the SSRN gets a high level of participation once authors know about it. Even though it is not the kind of prestige credentialing that is provided by peer review, there is a certain amount of validation. If your article has been downloaded 30,000 times, it is not just your sisters and cousins who are doing that, so this has some value to it.

Martin Blume added that the SSRN looks very much like the e-print server that is used in physics, and also like PubScience, a service that was provided by the Department of Energy until it was forced to be taken down. The APS mirrors the e-print archive and makes extensive use of it too, because authors can submit to that, and they subsequently submit to the society's home journals.

62

Authors' Assignment of Nonexclusive Licenses

Michael Jensen noted that it was not clear when he started out that he could make the SSRN work without taking anything other than a nonexclusive license, which people could revoke at any time. In fact, it does work. Indeed, he emphasized that he has never seen a publisher enforce a copyright in any way other than against an author. Whenever he gets a standard copyright assignment form from a publisher, he substitutes "non-exclusive license" for the word "copyright," and sends it back. Nobody has ever refused.

Digital media are making a difference, and attempts to maintain monopolies on the scientific literature are going to be very short-lived, perhaps another 10 years. In some areas it is going to be slower, and in others it will be faster, but licenses will be the future. Publishing certainly will not vanish but it will take a somewhat different form. Authors will have more rights. At the same time, publishers have to earn a fair return on their capital, or they cannot stay in business.

Ann Okerson agreed that the days of exclusive distribution may be rapidly fading. She recommended as a model the rights form used by the APS and others like it, based on a nonexclusive license.

She then read a question from a Webcast participant who asked: If the funding or employing institutions were to insist on retaining copyright of work submitted to STM journals, how would this impact STM publishers' decisions about acceptance, and also the overall economics of their publishing business? For example, the Mayo Clinic has in the past insisted on retaining rights after first publication, and this has had little impact on acceptance of papers originating from Mayo.

Martin Blume responded that the liberality of the APS's copyright form, which was mentioned, gives these rights to institutions and to authors. The society has a reason for insisting on taking copyright, as opposed to leaving it with the authors. Dr. Blume mentioned earlier that the society had put all of its content going back to 1893 online. If it had been taking during those years a license, and if that license had failed to mention the electronic means of distribution and publication, as it almost certainly would have, the society would have had to think quite a long time before going ahead, because it would have had to go to all the authors to collect these rights, at least if one were to believe the lawyers.

It is also not known how this will work in the future, for example, whether a license that gives the publisher rights to methods of distribution as yet uninvented is valid. APS has received legal advice that in a number of countries, this may not be a conscionable contract, and the society is a very international journal. Consequently, it would not be able to have the best assurance that it would be able to continue to assemble and distribute an archive in the future. This is the reason the APS asks for assignment of copyright. It does give back to authors all the rights that they would have if they had copyright, except the right to keep the society from doing what it wants to with the publication. It has not accepted a request for a nonexclusive license, and generally authors have understood this policy and complied.

Jane Ginsburg noted that if the license states that it extends to new modes of communication, now known or later developed, even though it is a license and not an assignment, that is valid. Even if you have an assignment of copyright, the author can terminate it under certain circumstances. She also noted that for those publishers who have any grant, exclusive or not, executed as of 1978, it can be terminated 35 years after the grant. Even with an assignment, therefore, a publisher does not have watertight protection, particularly for old material.

Dr. Blume then continued with the following story. I.I. Rabi, a distinguished professor of physics at Columbia, Nobel laureate, and one-time president of the APS, told an anecdote of taking President Eisenhower around the university. Eisenhower said what a pleasure it was to be able to meet with the employees of Columbia University. Professor Rabi drew himself up and replied, "We are not employees of Columbia University, we are Columbia University." Accordingly, perhaps the professors at Columbia could keep copyright under this, since they are the university.

Another reason the APS wants to hold copyright and have the maximum protection that it can get is that it does not trust the law or lawyers, not because lawyers are dishonest or the law is corrupt, but because there is no certainty about how any future dispute is going to come out. The *New York Times v. Tasini* case is a perfect instance of this, where an unexpected result came about. In fact, it is one that was detrimental to archiving everything on new media.

Although APS has never taken an author to court to enforce a copyright, it does deal with other publishers. It grants back all of the rights to the authors, but if other publishers are assembling collections of the society's articles and are going to make money out of it, the society needs to have the copyright to get a piece of the action.

With regard to the articles submitted by government employees, the situation is straightforward. Those articles are in the public domain, and APS accepts that. In fact, it could deal much more readily with works put in the public domain than with copyright being held by someone. It would be easier, because the society would then have a greater assurance that it would be able to do what it wished with the article in the future.

Nick Cozzarelli added that the policy of PNAS is the same as that of the APS. He went on to note that, as an author, he dislikes receiving requests for reproducing his work. He is delighted when a journal will do it for him, and will be even more delighted in the future when nobody has to ask. However, in the discussion thus far about the pros and cons of maintaining control of the copyright, nobody has brought up the fact that by the author maintaining sole rights, it is more difficult when someone else is trying to reproduce it and has to get permission from the author for it, instead of from the publisher.

Legal Issues in New Types of Digital Publications

Dan Atkins, of the University of Michigan, noted that subsequent panels at this symposium would focus on the future of the process and the products of publishing. In the future, the nature of the document could be expanded to be a multimedia work, involving computational programs, access to data sets, video, audio, and possibly produced as a composite object by multiple authors in multiple institutions, and so forth. What kind of complications does that present to the existing models of copyright and intellectual property?

Jane Ginsburg responded that the work that Professor Atkins described, a multimedia work, would probably be considered an audiovisual work under the copyright statute. The owners of that would be the creators, all of them, if it is not a work made for hire. This is a situation in which it might be useful to have contracts between all the owners before they go ahead with such a collaboration, because the administration of the rights to the work might be a little complicated.

The default position in the copyright law is that all joint authors share equally in the value of the work. Any joint author or co-owner of the work can separately grant nonexclusive rights in the work, unless the contract provides otherwise, although they have to get together and agree on the grant of exclusive rights. Then it is a question of appropriate contracting. As a matter of copyright law, however, it does not really pose any greater degree of complexity than making a movie or an audiovisual work with a lot of participants involved. It is largely a question of contracts.

Implications of Digital Rights Management (DRM) Technologies for STM Journal Publishing

Dan Atkins went on to ask about the DRM technology developments and the laws making it illegal to break the coding schemes that protect intellectual property rights. Do these developments have applicability to the STM realm, creating either additional opportunities or problems?

Jane Ginsburg responded that some DRM applications, such as pay-per-view publishable or downloadable could be applied to STM journals. There are technological measures that control access and possibly additional ones that control the copying of the work. If we just look at access control, for example, you might have a subscription deal that allows you to have access for a week. Copyright law prohibits you from breaking the access code, so that you could not keep it forever when you only paid to keep it for a week. Business models are now being developed on variations of how much of the work you get to see, how often, how many times, for how long, and so forth. All of these are built on access control technologies. The premise is that publishers can give people more varieties of access at different price points. This probably does not apply to STM as well as it does to the music business, where maybe you do not want to buy the CD with 12 songs you hate just to get the one song you want. There you can see that the model of accessing just one song with multiple replays could be quite attractive. It has taken the record companies a very long time to acknowledge that people do not want to be tied to the entire CD

64

anymore. They would like to have just one or two songs, regardless of whether they have it permanently or for some limited time. The package that the record companies have imposed on the music-buying public since the death of singles is now no longer enforceable, thanks to a variety of unlawful enterprises like Napster and Kazaa.

Similarly, one might say with respect to journal publishing, that not everybody wants the entire journal issue. Why do you have to buy the entire journal, if all you want is one article? The breaking up of journal issues has already been practiced in the print medium with respect to photocopy rights. The Copyright Clearance Center (CCC) has for years granted licenses for photocopying single journal articles. That translates perfectly well to the digital rights management situation as well.

An example of a publisher who enforced copyright in this context, not against an author, but against a user, was *American Geophysical Union v. Texaco*.[13] The American Geophysical Union sued Texaco because it was making systematically a large number of photocopies of works from that society and from 500 other journals, without paying for a license to do that. Texaco bought one subscription, and then circulated the subscription to all of its in-house scientists, who made and kept photocopies. Consequently, the American Geophysical Union, together with other publishers, sued Texaco, requiring the company to take the CCC license instead, and they prevailed.

Burdens for Small Publishers from Licensing and Restricting Access

Leslie Chan, of Bioline International, said that in many ways his project is the small deal compared to the big deal. Its associated publishers are all typically very small, nonprofit scholarly journals from developing countries. The major obstacle they have is making their materials visible, although by electronic means they are now able to do that.

Another obstacle has been getting their journals into libraries. They were advised a couple of years ago to take the licensing route and to negotiate with the libraries, and they quickly discovered that the big libraries have the advantage because they have the staff to negotiate a license. They also found the flip side to be true—that these processes are also set up in favor of the big publishers, because they have sales people and lawyers, people who do nothing but licensing and negotiation. Bioline International has a permanent staff of one and a half, so they just do not have the know-how or the time to negotiate with libraries or to work out a system to get their material into the right places. They are interested therefore in getting advice on how best to work with the library community.

Dr. Chan also noted that in the last year or so he realized that for small publishers with only several hundred members and very limited distribution, open access is in fact the only way to go. Not-for-profit publishers need to fulfill their mission to get the information and research out rather than to make a profit. There have to be other ways to recover the costs from other sources and activities, rather than getting stuck in the subscription model as a way to pay for the publication. A related lesson is that by controlling access, small publishers spend more money than they are able to recoup from the subscriptions. He would like to know from the other big publishers when they talk about the costs of IT, how much money they spend on controlling access.

Bioline International has 24 journals right now, some from Latin America, India, and Africa, and a couple from Southeast Asia. The number is growing, but the area that is growing the fastest is in fact the open-access journals, because, as they have been documenting in the last year and a half, the benefit of having the readership expand results in a lot more than just readership. It leads, for example, to submissions from authors outside their native countries, because the authors know that when they publish in these journals, which are peer reviewed, they will be read by their colleagues just the same. It is an aggregated database that will be Open Archives Initiative (OAI)-compliant.

Bioline International needs to measure the return in terms of the readership and the impact, rather than the revenue that it can generate. It hopes to go back to the funding agencies that support these journals, to convince them that they are spending the money in a good way

Martin Blume commented that the APS favors assistance to readers of the society's publications in developing countries, providing free access to them like many of the other societies. He agrees that for developing countries, in general, open access is the best approach. Moreover, there is an apparent dis-economy of scale between very small journals and very large ones, because the very large ones have all

---

[13] See http://www.arl.org/info/frn/copy/texaco.html

of the apparatus of employment, medical benefits, a human resources group, and so on. The smaller organizations do not have that.

One of the discussions that APS has had with a number of journals in developing countries is the possibility of their reprinting articles from the society's journals that were sent in by authors from their country, so that they can give greater publicity to those authors and also greater publicity to their journals. There are not that many articles that come from the smallest of the journals, but it would be worthwhile helping them with their publications.

Problems Encountered in the Transfers of Copyrights

Alan Rapoport, of the National Science Foundation, noted that, based on Jane Ginsburg's presentation, faculty may not even know who owns the copyright in the university. One can see the situation of faculty assigning the copyright to the publishers when they do not even own it. Does that happen often and what are the legal ramifications?

Jane Ginsburg responded that if an article is a work made for hire, then the author may in fact be selling the Brooklyn Bridge when signing the publishing contract. One way out of that bind is that a number of universities have their copyright policies in writing (see Table 5-1) and require their faculty to sign a special agreement, or their employment agreement incorporates the copyright policy by reference. If the university does not assert copyright, or grants back the copyright in traditional academic scholarship to the professors, that transfers the copyright to the professors, and then the professors have something to give to the publishers. It is true, however, that the ambiguity about the actual status of faculty writing potentially affects a lot of publishing contracts as well.

Problems with the University Work-for-Hire Approach to Academic Publications

Gordon Neavill, of Wayne State University, raised some more points in connection with the idea of academic work being done for hire. Many academics move from one institution to another. In that case, which institution owns the copyright? If he is teaching courses at Wayne State that he created earlier at the University of Alabama, it would be hard to see how Wayne State could claim copyright in lectures that originated elsewhere. The concept of academic freedom also suggests that the professor originates the concepts, develops them independently, and the work is not done at the behest of the university. All the university really requires is that the professor be productive, so how can the university claim copyright?

Jane Ginsburg responded that these are all very good reasons why academic work products should not be considered work made for hire. As she noted earlier, there has not actually been a court decision (with the exception of some inconclusive decisions in the Seventh Circuit Court of Appeals) about whether Congress perhaps unwittingly changed the law. The University of Michigan used to take the position that everything was work for hire, but has recently changed that.

66

6. WHAT IS PUBLISHING IN THE FUTURE?

INTRODUCTORY COMMENTS

Daniel Atkins, University of Michigan

The digital revolution is disaggregating the traditional processes of many knowledge-intensive activities, but in particular the publishing or scholarly communication processes, and it is offering alternatives in both how the various stages of these processes are conducted and who does them. The various functions—whether metadata creation, credentialing review, or long-term stewardship—can be separated, disaggregated, and players different from those who traditionally have carried out these tasks can, in theory, perform them.

The digital revolution is also changing the process by which knowledge is created, by which discovery takes place. This is most true within the scientific, technical, and medical (STM) arena. That is the central theme of a National Science Foundation (NSF) study, *Revolutionizing Science and Engineering Through Cyberinfrastructure*, recently chaired by Professor Atkins.[14] The report clearly documented an emergent vision and enhanced aspiration of many science communities in the use of information and computing technologies (ICT) to build comprehensive information environments based on what is now being called cyberinfrastructure, such as collaboratories or grid communities. These are functionally complete, in the sense that all of the people, data, information, and instruments that one needs to do a particular activity in that particular scientific community of practice are available through the network and online.

A growing number of research communities are now creating ICT-based environments or cyberinfrastructure-based environments that link together these elements with relaxed constraints on distance, time, and boundaries. There is a general trend toward more interdisciplinary work and broader collaborations in many fields.

Publications now exist in many intermediate forms. We are moving toward more of a continuous-flow model, rather than a discrete-batch model. Raw data, processed data, replays of experiments, and deliberations that are mediated through a collaboratory can be captured, replayed, and re-experienced. Working reports, preprint manuscripts, credentialed or branded documents, or even post-peer-review annotated documents can now become available at varying times to different people with diverse terms and conditions.

Publications need not necessarily be precredentialed before publication on the Net. As George Furnas, at the University of Michigan, says, their use on the Net can itself be credentialing. In theory, every encounter with the document may be an opportunity to rank it in some way and to create some kind of a cumulative sense of its impact or importance. There could be alternative credentialing entities or methods, and you could pick your favorite.

The raw ingredients—the data, the computational models, the outputs of instruments, the records of deliberation—could be online and accessible by others, and could conceivably be used to validate or reproduce results at a deeper level than traditionally has been possible. The primary source data can be made available with a minimum set of metadata and terms and conditions. Third parties—particularly in an open-access, open-archives context—can then add value by harvesting, enriching, federating, linking, and mining selected content from such collections.

---

[14] National Science Foundation. 2003. *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. National Science Foundation, Arlington, VA.

The goal of this session is to try to illuminate and inform the discussion about some of these emerging technologies, the related social processes, some specific pilot projects, and the challenges and opportunities that may provide the basis for these kind of future "publishing processes." This is put in quotes, because we may someday not actually think about it explicitly as a publishing process, but as more holistically integrated into the "knowledge creation process."

IMPLICATIONS OF EMERGING RECOMMENDER AND REPUTATION SYSTEMS

Paul Resnick, University of Michigan

When some people think about changing the current publication process to a more open system, they express concerns that scholarly communication will descend into chaos, and that no one will know what documents are worth reading because we will not have the current peer-review process. This idea should be turned on its head. Instead of going without evaluation, there is a potential to have much more evaluation than we currently have in the peer-review process. We can look at what is happening outside of the scientific publication realm on the Internet to give us some clues about where this could go.

The Democratization of Review and Feedback

In today's publication system, there are reputations for publication venues. Certain journals have a better reputation than others, and certain academic presses have strong reputations. There are a few designated reviewers for each article, and these serve as gatekeepers for the publication. An article either gets into this prestigious publication, or not; it is a binary decision. Then afterward, we have citations as a behavioral metric of how influential the document was.

We can examine some trends on the Internet to see how they apply to the scientific publication and communication process. There can be a great deal of public feedback, both before and after whatever is marked as the official publication time, and we can have lots of behavior indicators, not just the citation counts.

Let us consider some examples of publicly visible feedback. Many Web sites now evaluate different types of products or services, and post reviews by individual customers. In the publishing world, many people are now familiar with the reviews at Amazon.com, both text reviews and numeric ratings that any reader can put in. Many of us find this quite helpful in purchasing books. We do not quite have it for individual articles in scientific publishing yet, but we do have it for books, even some scientific ones.

Even closer to the scientific publishing world, there is a site called Merlot, which collects teaching resources and provides a peer-review process for them. There is a peer-review process before a publication is included in the collection, but even after it is included, members can add comments. Typically, such comments are made by teachers who have tried using it, and they report what happened in their classroom, and so on. The member comments do not always agree exactly with the peer-review comments.

These examples provide a sense of what is happening with subjective feedback that people can give on the Internet, beyond the traditional peer-review approach.

Behavioral Indicators

With behavioral indicators, you do not ask people what they think about something, you watch what they do with it. That is like the citation count. For example, Amazon.com, in addition to its customer reviews, has a sales rank for each book.

Another example is Netscan, which is a project that Mark Smith at Microsoft Research has been doing to collect behavioral metrics on Usenet newsgroups. It uses various types of metrics about the groups.

Google uses behavioral metrics of links in their page rank algorithm. Many people check how they are ranked on Google on various search strings. However, Google is not just doing a text match; it

68

also takes into account how many links there are to the page from other Web pages, and weights it by the rank of the other pages. So this is a ranking system that takes into account the behavioral metric of who is linking to whom.

A final example is the Social Science Research Network that was described by Michael Jensen. It uses the download count as a behavioral metric. The main one is the all-time, "top 10" downloads, but it also has many other different top-10 categories, so that more of the authors using the system have a chance to be a winner.

Issues in Evolving Credentialing Processes

The preceding examples provide some potential models for developing credentialing processes for scientific publication in the future. There are also some problems that require further thought. Some of these already are active areas of research for people who are working on recommender and reputation systems.

An obvious potential problem is gaming the system. You can make a bunch of Web pages that all point to yours so that Google will rank yours higher. In fact, there is a whole cottage industry that does this. You can hire consultants who will help you get higher in the Google rankings.

It is a little harder to do this with the Amazon sales rank. It requires you to actually buy some books. No matter what the system, however, people try to figure out what the scoring metric is and game the system. This means that those who are designing these metrics need to consider it from the outset. The ideal metric would be strategy-proof, meaning that the optimal behavior for the participants would be just to do their normal thing and not try to game the system, but it is not always so easy to design the metrics that way.

Another problem is eliciting early evaluations. In those systems where there is widespread sharing of evaluations, there really is an advantage to going second, to let somebody else figure out whether an article is a good one to read or not. Of course, if we all try to go second, there is no one who goes first.

Yet, another problem can be herding, where the later evaluators do not really reveal what they thought of the document, but are overly influenced by what the previous evaluators thought—they just go along with the herd. There are some interesting approaches that potentially would help with the herding problem. For example, you might reward evaluators for saying something that goes against the previous wisdom, but with which subsequent evaluators agree. That would be the person who finds the diamond in the rough and that person would get special rewards; the person who just gives random or tendentious reviews would get noticed and would get a bad rating. Such a process also would require going back and revisiting some of the decisions that we have made about anonymity and accountability in review processes: single-blind, double-blind, not blind at all. For different purposes different processes may be preferred.

Experiments to Try in STM Publishing

Some potential experiments are more radical than others. Journal Web sites might publish reviewer comments. The reviewers might take more care if they knew their comments were going to be published, even without their names attached. The reviews for the rejected articles could be published as well. There could be fewer really bad submissions if authors knew that they could potentially be hurting their reputation by having the reviews of their article up there. Then, after publication, the Web sites that publish full-text articles or abstracts could let people put comments there that would be publicly visible.

Some other experiments might try to gather more metrics. Projects such as CiteSeer in the computer science area measure citations in real time. One might also use the link and download to find out how many people are actually reading online, how many times an article is being assigned in courses, and other behavioral metrics.

Experiments in evaluating the evaluators are needed as well. The best place for this might be in some of the conference proceedings, where there are a number of evaluators for an individual article. One could examine the nature of the reviews in a more explicit attempt at evaluating evaluators.

More attention, greater credit, and rewards need to be given to reviewers for evaluating early,

69

often, and well. Publishers already are complaining about the difficulty of finding good reviewers. One reason is that reviewers' services are not valued sufficiently in the system. We need some metrics that go beyond merely noting that someone was a reviewer or an associate editor at some journal, but that actually evaluate and give proper credit to such functions. We might start thinking about that as a contribution to research, rather than as purely a service line. In the promotion and tenure reviews for academics, the categories typically include teaching, research, and service, but this evaluation and commentary activity might really be thought of as contributing to the growth of knowledge. If some metrics could be developed on how much they are contributing in what way, we might think about that as a research contribution rather than just a service contribution.

PREPRINT SERVERS AND EXTENSIONS TO OTHER FIELDS

Richard Luce, Los Alamos National Laboratory

The preprint service is something that is a well-known, well-understood concept in the physics community, but not much in other communities. It is useful, at the outset, to distinguish between preprints and e-prints. Preprints have a "buyer beware" connotation to them in the physics community. They provide a means to obtain informal, non-peer-review feedback that is weighted very differently in the community than a formal refereed report. They are a way to get an early version of an article out to colleagues to get some feedback, if any may come back, to help decide whether or not to publish it later.

e-Prints, on the other hand, tend to be more polished papers deposited by authors in a public online archive in order to speed up the communication process and give authors more control over the public availability of their work.

The e-Print arXiv in Physics

Back in 1991, Paul Ginsparg created the e-Print arXiv[15] at the Los Alamos National Laboratory. That database archive, which today is at Cornell, has about 30 or so fields and subfields and 244,000 papers. It has succeeded in large part because Dr. Ginsparg is a physicist. As a high-energy physicist, he understood well how that community works and what its needs are. His notion was to take and streamline the communication process relative to preprints.

Over the past decade, this approach has spread to some other fields, including mathematics, materials, nonlinear sciences, and computation. The adoption of this approach by other disciplines demonstrates that this is not a phenomenon only in high-energy physics or in the physics community, but can work in other research areas.

It clearly has increased communication in the areas that it covers in physics. It is the dominant method for authors to register publicly when their idea first comes out, even though the article may not be formally published until six months later.

According to Dr. Ginsparg, the cost is very low, and this is partially responsible for the wide acceptance of the system. There has been a continuing increase in submissions. The driver in the community clearly is speed, to make communication faster. The e-Print arXiv has set an example for both society and commercial publishers to consider. It is significant to note that in 1995-1996, the American Physical Society (APS) began to accept preprint postings, and later began to link back to them. This was the beginning of a more formalized recognition that there was a role for this approach—a bottom layer or first tier of scientific information—and that we could have a two-tier structure and start to link those things together.

One might raise the question of the quality of the information in the e-Print arXiv. These articles are not peer reviewed. If one does an analysis over a period of time of the quality of the submissions, however, what you see is a field-specific track record in terms of what actually gets published.

In high-energy physics theory, about 73 percent of the papers in the archive turn out to be

---

[15] For additional information on the e-Print arXiv, see http://www.arXiv.org/.

70

published, whereas in condensed matter physics, it is only somewhere around one-third. Consequently, it is fairly field specific, but it does indicate that the archive does not just contain articles that have no future or role in the formal publication system itself.

What lessons can we learn from this? Greater timeliness in the disclosure of research results is certainly one. Another is that a few passionate people can make a difference. For a decade, while this system was at Los Alamos, it was typically run by three to five people, sometimes spending far too many hours, but inspired by the view that this was really going to change the world. That small number of people and the relatively small amount of money to operate the system (about a half million dollars per year) became a very dominant factor in terms of the community itself. Most important is the lesson is that it addressed the common interests of a community in a sociologically compatible way, which is why the system worked for that particular community.

Other Preprint Server Projects

Scholarly communication is a very complex ecosystem. Clearly, not all fields are the same. The sociology, behavior, and traditions differ from field to field. Consequently, this solution is not the correct or only solution, nor could it be expected to fit in all other fields. One needs to understand the community behavior and traditions and then look for models that meet those kinds of needs and requirements.

Nonetheless, there have been spinoffs into other fields organized by both universities and government agencies. CogPrints, at the University of Southampton in the United Kingdom,[16] is a preprint system that is quite well known in cognitive science. NCSTRL, the Networked Computer Science Technical Reference Library, was an early effort to get computer science papers harvested together and then start to build a federated collection of that literature.[17] NTLTD, developed by Ed Fox at the University of Virginia, provides thesis dissertations.

Within the federal government, the NASA National Technical Reports Server (NTRS)[18] was a pioneer in terms of trying to bring together and to make available a collection of federal reports, both in terms of metadata and the full text. PubMed Central is certainly well known in the life sciences community. *Living Reviews*, (www.livingreviews.org/) is a somewhat different model in gravitational physics at one of Max Planck's institutes in Germany, created a review that gets updated by the author over time. Rather than going back to read a decade-old review that is static and wondering what has happened to the field since it was published, readers of *Living Reviews* know that it is current, because the authors have committed to a process of keeping the material they put into that online publication up to date.

There are perhaps a dozen well known e-print systems with servers active today. There are around 100 other servers who claim to have some kind of an e-print system. They use a standardized protocol and allow people to come in and at least harvest some of the material that is collected there.

One problem is the absence of an enabling infrastructure. The open archives initiative was not meant to be the end-all, be-all fix to the system. It sought a solution that would allow a discipline-specific e-print archive to be able to talk to or communicate with other systems, so that users would have the opportunity to look at a pool of material. The problem is, how to ensure access to this variety of different systems.

The protocol specifies the method by which material on the different systems can be harvested. We are just reaching the point where we are starting to see what kinds of interesting services can be put on top of that. That development perhaps has been slower than expected, but is now beginning to take off with a variety of different systems. One example is CiteSeer, mentioned by Paul Resnick. This begins to hint at the kinds of things that people might do in an open environment.

Implications for the Future

What do these developments mean beyond the physics community, and what new efforts can we

---

[16] For additional information on CogPrints cognitive sciences e-print archive, see http://cogprints.ecs.soton.ac.uk/.

[17] For additional information on the NCSTRL archive, see http://www.ncstrl.org/.

[18] For additional information on NASA's National Technical Reports Server, see http://ntrs.nasa.gov/.

71

foresee? It is important to note that there has been very powerful opposition coming from traditional institutions that have used journal publishing as a cash cow, or from secondary publishers who see their secondary databases essentially as a birthright. For example, there was a lot of political pressure leading to the demise of the Department of Energy's PubScience Web portal because it was perceived as threatening to those other kinds of interests.

A major question that needs to be addressed is how peer review will work in the preprint and open-access context. How should we evaluate, recognize, and reward? Another question that is related in part to the peer review issue is: What is the influence, and how do we detect it?

There are a variety of methods that one needs to consider, essentially a composite approach. Today we use citations as the sole indicator of influence, that is, an author-derived statement about what is important, what has influenced the research.

We might look at a complementary path, which is the notion of reader behavior related to determining influence. Digital libraries or service providers can provide analytical tools to generate new metrics based on user behavior, which complements or may even surpass citation ranking and impact factors.

What is the problem with impact factors? To some extent, it is the lazy person's notion of how to figure out what is important in terms of journal ranking. It can be very convenient for publishers to state that their journal is ranked well in terms of impact factors. It also is relatively easy for librarians to justify why they buy one title over another title.

The problem is that the citation is only an indicator of influence. There are many reasons that people might cite a paper: to show that they have read the literature in a field, to disagree with somebody and prove them wrong, to help a colleague get some visibility and enhance reputation, or to give credit to someone's ideas. Although impact factors are widely used to rank and evaluate journals they frequently may be used inappropriately. Then there is a whole field of bibliometrics, which enables one to track science and to find out the interrelationships between authors, citations, journal citations, and the subjects that are covered. This is still an emerging field, but one that may become more extensively used.

A better approach, however, might be to supplement the current system with a multidimensional model to balance bias. What would such a model look like? An ideal system might have the following elements: citations and co-citations to determine the proximity indicator; the semantics, or the content and the meaning of the content in articles to see how they are related; and user behavior, with regard to behavior metrics. At Los Alamos, about 95 percent of information resources are electronic-only, so it is possible to detect and determine community-specific research trends and look at where those trends differ from the ISI impact factors. However, this is quite complex and raises privacy concerns.

Finally, there is the problem of long-term curation, which has several aspects. The first element is the published literature, which is what most people think about in the context of preservation. There is also the issue of the relationships in the rich linking environment that one might want to collect and preserve over time, however, and make available in the future.

INSTITUTIONAL REPOSITORIES

Hal Abelson, Massachusetts Institute of Technology

This panel is supposed to be about this wonderful cyberinfrastructure future. One is reminded of William Gibson, the outstanding cyberpunk writer, who 20 years ago gave us the word cyberspace. He said that he does not like to write about the future, because for most people the present is already quite terrifying enough. It is in that spirit that this presentation looks at the present.

The Changing Role of Universities in Making Their Research Output Openly Accessible

The main action one should watch for is not in the new technology, but the possibility of new players finding institutional reasons related to their other primary missions to participate in a new game of disintermediate thy neighbor. In particular, do universities have institutional roles to play here other than what they have done so far, which is to be the place where the authors are? Do universities have a

reason in their institutional missions to start participating in making their research output openly accessible?

According to MIT's mission statement (and probably that of many other universities), MIT is committed not only to generating knowledge, but also to disseminating and preserving it. So how does a place like MIT implement this mission? One initiative, OpenCourseWare,[19] is already getting quite famous. According to the MIT President's Report in 2001, the university has made an institutional commitment "to provide free access to the primary materials for virtually all our courses. We are going to make our educational material available to students, faculty, and other learners, anywhere in the world, at any time, for free."   The reason MIT did that is not that it was overcome by some fit of altruism; rather, it decided that given the way the world is going, it would be better for MIT and indeed all universities, in terms of fulfilling their primary mission to educate students, to put their primary educational material on the Web.

The OpenCourseWare Web site is a prototype and has 50 courses up. There is a group at MIT that is madly trying to get the first 500 courses online by September 2003. They are on a timeline to get all MIT courses up by 2007. It is an institutional publication process, to which the university has committed as a permanent activity.

The DSpace initiative,[20] which is the sister project of OpenCourseWare, is a prepublication archive for MIT's research. The difference between DSpace and most other prepublication activities is that there is an institutional commitment by the MIT libraries, justified by MIT's mission to maintain it.

OpenCourseWare would make sense if only MIT did it, but DSpace cannot possibly be like that. DSpace, in addition to being a preprint archive for MIT, is meant to be a federation that collects together the intellectual output of the world's leading researchers.   MIT now has six institutional partners working with it. What is important about DSpace is that there is a group of universities working out the management and sustainability processes in terms of their institutional commitments for how to set up such a federated archive.

Responses to the Proprietization of Knowledge

Both OpenCourseWare and DSpace are ways that MIT and other universities are asking what their institutional role should be in disseminating and preserving their research output. Why are these questions coming up now? Why might universities start wanting to play institutional roles in the publication process, other than serving as places where the authors happen to be?

The answer is that the increasing tendency to proprietize knowledge, to view the output of research as intellectual property, is hostile to traditional academic values. What are some of the challenges that universities see? They include high and increasing costs; imposition of arbitrary and inconsistent rules that restrict access and use; impediments to new tools for scholarly research; and risk of monopoly ownership and control of the scientific literature.

The basic deal, as seen by universities, is that the authors, the scientists, give their property away to the journals. The journals now own this property and all rights to it forever. Lifetime of the author plus 70 years is forever in science. If that regime had been in place 100 years ago, we today would be looking forward to the opportunity in 2007 to get open access to Rutherford's writings on his discovery of the atomic nucleus. Then the publishers take their property and magnanimously grant back to the authors some limited rights that are determined arbitrarily and totally at the discretion of the publisher. The universities, who might think they had something to do with this, generally get no specific rights at all, and the public is not even in this discussion.

Jane Ginsburg already showed some examples of the restrictions imposed by publishers on authors, but it is most useful to mention some here. For example, some rights generously granted by Reed Elsevier to its authors include the right to photocopy or make electronic copies of the article for their own personal use, the right to post the final unpublished version of the article on a secure network (not accessible to the public) within the author's institution, and the right to present the paper at a meeting or conference and to hand copies of the paper to the delegates attending the meeting. The rights

---

[19] For additional information on MIT's OpenCourseWare project, see http://ocw.mit.edu/index.html.

[20] For additional information on the DSpace Federation, see http://www.dspace.org/.

73

generously granted to authors by the journals of the American Chemical Society are that authors may distribute or transmit their own paper to not more than 50 colleagues, and authors may post the title, abstract (no other text), tables, and figures from their own papers on their own Web sites. The rights from the *New England Journal of Medicine* to its authors are even more limited: "The (Massachusetts Medical) Society and its licensees have the right to use, reproduce, transmit, derive works from, publish, and distribute the contribution, in the Journal or otherwise, in any form or medium. Authors will not use or authorize the use of the contribution without the Society's written consent, except as may be allowed by U.S. fair-use law."

It is instructive to list some of the elements that are valuable for promoting the progress of science. They include quality publications and a publication process with integrity, certainly, but also: open, extensible indexes of publications; automatic extraction of relevant selections from publications; automatic compilation of publication fragments; static and dynamic links among publications, publication fragments, and primary data; data mining across multiple publications; automatic linking of publications to visualization tools; integration into the semantic web; and hundreds of things no one has thought of yet.

Will the information technology to support scholarly research be stillborn because everything is hidden behind legal and electronic fences? Probably not, because it is too valuable and people are going to invest in it anyway. The more serious possibility is that the spread of these tools will be done in a way that stimulates network effects that will further concentrate and monopolize ownership of the scientific literature. If a search engine only searches the publications of one publisher, that becomes valuable enough for the publisher then to come in and do what the librarians call the big deal. As Derk Haank, CEO of Elsevier Science, noted recently: "We aim to give scientists desktop access to all the information they need, for a reasonable price, and to ensure that the value of the content and the context in which it is presented are reflected in the information provision. The information is made available to researchers under licenses accorded to their institutes, and they have all the access they wish."

Does this mean that science will be restricted with monopoly ownership, or rather will we enjoy a system that participates through open standards? One impediment to openness is copyright. It turns out to be surprisingly difficult—in the wonderful legal phrase—to abandon your work to the public domain. It is even more difficult to specify some rights reserved, rather than all rights reserved, which is the default rule under copyright. The Creative Commons was founded recently to encourage people to allow controlled sharing of their work on the Internet.[21]

To sum up, the world is disaggregating, and there is a big game of disintermediation going on. The place to look for access is not new technology. The places to look for the action are the new institutional players coming into the game. New technologies might enable new roles for the universities, and that will lead to the promotion of the progress of science.

DISCUSSION OF ISSUES

The Role of Peer Review and Other Review Mechanisms

Steven Berry began by pointing out that there is one aspect of the refereeing process that Paul Resnick either dismissed or overlooked. Reviews by and large have a lot of influence on what actually gets published. It is not a simple rejection process. Of the papers that are reviewed and published, a very high percentage are revised as a result of the reviews.

Furthermore, we have to recognize that the review process provides only a very low threshold. It simply identifies material that is of sufficient quality for scientific discourse. So it is not necessarily a judgment of whether it is right or wrong. This function could be done in other ways, of course, but we have to recognize that in some fields, reviewing even at that low threshold is looked upon as a very important protection.

The arguments between physical and biological scientists about this are instructive. Physical scientists, as Rick Luce pointed out, are very ready to accept the archive model without review first, and use online reader review. The biological scientists are worried that without that low threshold review, information could be put online that might be dangerous to lay users. They feel that there is a large

---

[21] See http://www.creativecommons.org.

74

audience of nonprofessional readers of biological, and especially biomedical, articles that simply does not have the judgment that the professionals have, and that it is simply dangerous.

Paul Resnick responded by emphasizing that in considering these alternative review mechanisms, he views them as being in addition to peer review, not instead of it. We also could get more out of the peer-review process than we are getting now, if the reviews were made public. For example, you do not want to let bad information get out there for the general public to see without anybody from the scientific community saying it is rubbish. Fine, put it out there along with the reviews of scientists who said it is rubbish. Why is it better to just hide it than to disclose it with the commentary from the scientific experts?

Dan Atkins read a question from a member of the Webcast audience, who asked: Might not the thread of the original version of a paper, along with reviewers' comments and authors' revisions or responses to the comment, as well as the journal editor's additional comments and interpretations, be used as educational material and enrichment for university students? This could be done either anonymously or by attribution.

Paul Resnick said he thought it was a great idea. One of the problems for Ph.D. students is that, more often than not, they do not get to see a paper through its whole process until they do it themselves and have cleared it the hard way. An adviser can show them the reviews submitted on the adviser's papers, but this is not done routinely. Making it a more public process would be helpful for education.

Rick Luce added that there needs to be a proper balance in that discussion, whether we should let the system filter out the bad material or let the user filter it out.

The Role of the Public Press

Ted Campion noted that one big player in scientific publishing that has barely been mentioned is the public press. Scientific publication, particularly in biomedicine, is largely being judged now, at least by authors and even academe more generally, by how much press coverage it gets. It is not just studies of estrogen. Zebrafish and hedgehog mutations are getting into the press. The scientific press is not only being judged now by the boring citation indices, but by whether the network news covers it. This, of course, is all being driven by the public's increasingly sophisticated understanding and interest in science, and in biomedical sciences in particular. What effect is this having on scientific and biomedical publication?

Hal Abelson responded that he had a discussion with his daughter, who is in medical school, about MIT's OpenCourseWare, and she told him she thought it would be a terrible thing for medical schools to put their course curriculum on the Web, simply because you had to be a professional medical student in order to evaluate and use that information, so it would be dangerous to have it out there.

It is hard to know what to do about sensationalism in the public media. At the same time, you could do a lot worse than to have Peter Jennings talk about an article in the *New England Journal of Medicine*. It has been a tradition in the United States that the cure for speech is more speech. Maybe if there were other channels for people to respond, things would be better, but restrictions on the press are not the answer. Maybe part of the answer is that it is up to the press to worry about the novelty and up to the journals to worry about the authenticity.

Curation and Preservation of Digital Materials

Mark Doyle, of the APS, said that like MIT, part of his society's mission is explicitly the advancement in the diffusion of knowledge in physics. In addition to peer review, the society considers very important the responsibility to do the archiving. It has already gone through the transition, going to fully electronic publications. The core of its archive is a very richly marked up electronic file from which the print and the online material are all derived. What still appears to be absent in DSpace or in the e-Print arXiv are any efforts to build the infrastructure for curating that kind of material. It is one of the most important things to do.

One other related issue is the low cost of arXiv.org. The key problem is that there is a two to three order of magnitude difference between what it costs the society or other publishers to publish—$1,500 per article—and what it costs to publish in the arXiv. That is really where all the tension in the

economic models come from, when you have such large orders of magnitude difference. That is what puts all the pressure on people to change the way that things are done.

Rich Luce responded that curation is an interesting question. Most libraries are not going to be able to do this. There are probably one to two dozen libraries globally that see that as an important role, together with a significant capability in which to make an investment, because they are thinking about centuries of preservation, not years or decades. There are many publishers who are quite aware of the problem, but very few make the necessary investments. The vast majority of publishers are simply too small to have the wherewithal, both technologically and financially, to be able to pull it off. It is going to take some sort of a hybrid relationship between some publishers and some libraries, who see that as their role.

Hal Abelson added that when MIT designed the DSpace, it was absolutely essential and deliberate that it be housed in the MIT libraries. The reason was that whether or not the MIT libraries will preserve something for 200 years, they are certain to preserve it for 50 years. The developers of DSpace wanted to work with an organization that understood what archiving and curation meant, and that is what libraries do. To build on what he said earlier, the critical thing is not the technology for archiving, because that technology is well under control. Rather, it is to find an institution that will commit, as part of its core mission, to keep and preserve the material. For example, if it is part of the core mission of the American Chemical Society to be the repository of all chemical literature and to have every other organization in the world be its franchisee, that is an important thing to say.

It also is very important not to get trapped into the idea of building the monolithic, end-to-end solution. DSpace will never be that. It might be a place where people who are building peer-review and curating systems, and all kinds of other systems, can link to and build on. The trap, which was alluded to yesterday by Gordon Tibbitts, is that you do not want to be in a position where you build the whole system. Instead, you want to have elements that are communicating through open standards, so that lots of people can come in at different places in the value chain and add value in different ways.

Evaluating STM Literature Outside Academia

Donald King noted that in his studies of the amount and types of use of STM articles, approximately 25 percent of those articles are read by academicians, and the rest are read outside of the academic community. When you begin these feedback systems that Paul Resnick was describing, you need to think in terms of the enormous value that is derived from the literature outside academia. There are two purposes for doing this. One is that it is a better metric for assessing journals and authors. The other is that it also will begin to develop a means for the authors to better recognize that their larger audience is outside of their immediate peers.

Professor King added that he has done a lot of focus group interviews and in-depth interviews of faculty, and they seem to think that they are writing only for the people they know, their immediate community. There would be some value in the assessment system, therefore, if there were some acknowledgment or recognition that there are other uses of that information outside academia.

Paul Resnick agreed that providing such external feedback to the authors on how their works are being used would be useful.

Adapting Tenure Requirements to Open Source

Lennie Rhine, University of Florida, asked how universities adapt the tenure process to the open-source environment. Most academics are tied to the peer-reviewed journal system as a mechanism to be ranked hierarchically and to be evaluated in the tenure process. How do you incorporate this more ephemeral open-source information into that process?

Paul Resnick responded that in the current tenure process, the department heads look at the journal rankings and count the number of journals in which you are published and cited, and different departments weight it differently. They may or may not construct a numeric score. One could actually develop a more open system for computing metrics like that. Consider how *U.S. News & World Report* does its rankings of schools; it has a particular way of weighting everything. Now, imagine a more open version where we collect all the data, know what things have been cited and read, have all the reviews,

76

and have both the behavioral and the subjective feedback. Then the teaching institution can create its own metrics based on its own tenure criteria, and the research institution can create a different metric, and you can have many different ways of using those data.

Hal Abelson added that one of the marvelous things that the NSF did several years ago was to limit the number of papers that one can cite in preparing a grant proposal. It is not really a question of quantity or vast numbers of papers published in third-rate journals. MIT has been trying to get its teaching staff who are coming up for tenure to identify three things on which they should be evaluated. The goal is to try to get the enterprise focused on what Alan Kaye used to call "the metric of Sistine Chapel ceilings per century, rather than papers per week."

Problems with Metrics

Martin Blume was happy to hear Rick Luce use the word indicators, rather than metrics, because there is no one number that can be used as a measure of quality. There are things wrong with all of them, including the opportunities for gaming the system.

There is a *Dilbert* cartoon that shows the human resources manager saying that metrics are very important, and a very good one is the rate of employee turnover. The manager replies that they do not have any turnover, they only hire people who could not possibly get work anywhere else.

Many metrics suffer from this, and they can be manipulated. You really have to look into them and use them as indicators, and it takes a fair number of them if you are going to get a fair measure of quality.

Open Versus Confidential Peer Review

Dr. Blume also commented on peer review, and the concerns about public comment. Although he thinks public commentary is a good thing to do, nevertheless there is a sort of Gresham's law of refereeing, in that the bad referees tend to drive out the good ones. All of us who take part in listservs of one sort or another know of the loudmouth who will not be contradicted or denied, and eventually the rest of us give up and say we are not going to take part in this anymore. You need to expect something like this, and you have to have a degree of moderation in it. That is where an editor's role in the peer-review process comes in. Also, the knowledge that a paper is going to be peer reviewed does have an effect on authors. It means that they try to improve it at first so that it will pass this barrier.

Dr. Blume then presented some statistics on peer reviews from the APS's journals. The society looks at the first 100 articles submitted to one of its journals and tracks them through a year to see what has happened to them in the end. Of the first 100 submitted in one of the years, 61 were accepted and the remaining 39 were rejected or recommended for publication elsewhere. Of the 61 that were accepted, 14 were approved without any revision after one report, 22 were approved after resubmission after some modifications, 14 after a second review, and 4 after the third resubmission, all of these leading to improvements. Some wags would say that the improvements for some of them are largely adding references to the referees' papers, but even that is an improvement if there is not enough citation of other work. Of the rejects, 19 were rejected after one report, 13 after two, 4 after three, 2 after four, and 1 after six reports. It is much more costly and difficult to reject a paper than to accept it.

This provides an indication of some of the value added in the course of traditional peer review.   It is something that has to coexist with the other types of assessments.   APS also tries to avoid using referees of the type that would lead to Gresham's law such as one sees on the listserv. It is aware of them, and tries to select accordingly.

Paul Resnick said that one way to evaluate the reviewers is to have an editor who chooses them or moderates. You also could come up with some system where you calibrate reviewers against each other.

Dr. Blume noted that his society actually does this. It keeps a private file based on the reports it receives. Unfortunately, this leads to an overburdening of the good reviewers, so they are punished for the good work that they do.

Dr. Resnick pointed out that that kind of system that Dr. Blume is using privately and internally could be adapted to a more public version. If you go to a more public system, you do not necessarily

have to give all the lousy reviewers an equal voice.

Dr. Blume added that his society would certainly want to pick people and would probably continue to do it anonymously. A reviewer is always free to reveal his or her identity, but the APS will not.

Relationship Between Open Archives and Traditional Publishers

Fred Friend asked Hal Abelson how he sees the future long-term relationship between open repositories and traditional publishers. One can see that repositories such as DSpace have a very valuable role in shaking up the system and in helping establish or return to better priorities in scholarly publication, but what is the future role for traditional publishers in that situation? Hopefully, publishers will respond in a positive way to these changes, and may come out in the end with a better role than they have at the moment. Or could institutional repositories take the place of traditional publishers completely?

Professor Abelson noted that it was Yogi Berra who said, "It is really hard to make predictions, especially about the future." The main point is that as you have new players, you have different kinds of roles. There is no inherent hostility between institutional archives and traditional publishers. MIT, for example, has a very respectable journal operation in the MIT press, and it is looking for ways to find joint projects with the DSpace archives. One can imagine a university holding both the preprint through the edited version, and the journals coming in with some kind of authentication and review cycle. There may be lots of opportunities. The trick is to free up the system and allow other players to provide pieces of that process that the journals, for various reasons, have not been.

The danger is for some individual player to come in and try to lock everything up in some complete system. The problem with the World Wide Web has always been that everybody wants to be the spider at the middle of it, and that is the outcome that we have to resist.

Dan Atkins recalled the point made by Rick Luce that preprint servers are the repositories at the lower layer, and they provide a platform or an infrastructure on which a whole host of yet to be fully imagined value-added entities could be built, some of them for-profit entities. The idea is to create a more open environment for the primary, or upstream, parts of the value chain, and then to encourage an economy of activity on top of that.

One of the themes that came out loud and clear in the NSF's recent cyberinfrastructure study was the huge latent demand for serious curation of scientific data, and mechanisms for the credentialing and validating of data and for encouraging interoperability between data in different fields as people create more comprehensive computational models. Of course, there is a lot of synergy between that and some of the long-term preservation and access issues. In fact, the volume of bits that, for example, just the high-energy physics community generates in a year probably exceeds most of the scientific literature worth keeping.

Fred Friend added that, traditionally, the formal publications have been viewed as being the record. Yet we seem to be saying that long-term archiving is not for publishers. So that perhaps rules out the recordkeeping function for traditional publishers. What are they then left with?

Rich Luce noted that an issue of deep concern to the governmental sector, and to public institutional repositories, is being able to have access to material created with public monies, and to make these materials publicly available. One can easily imagine a system again that is open at the bottom level, where there are some nuggets that perhaps publishers might look at and start to mine for opportunities to add value to on a more formal basis. These activities do not have to be competitive. In fact, they can coexist in a way that is very complementary.

Universal Search Engines for Licensed Proprietary Content

Mark Krellenstein, from Reed Elsevier, referred to Hal Abelson's comments about Derk Haank's statement about Reed Elsevier producing a universal search engine for its licensed and proprietary content. That initiative came in response to the publisher's users and the libraries. Reed Elsevier has found that people want as close to a universal search engine for proprietary materials as they have with Google for the open materials on the Internet.

The idea of multiple players in the value chain, which appeals to expert researchers, frequently is

less appealing to undergraduates, in particular, who really want a single solution for all their search needs. Reed Elsevier does not expect to be that single solution, but it is trying to do as much as it can. It is charging for it because of the model it has for charging for content. Google supports itself via advertising. That could be an option perhaps for a company like Reed Elsevier, but probably not in the scientific context to make the kind of revenue that is necessary, given other models discussed at this symposium.

Another point is that Reed Elsevier is also open to other players doing the same kind of thing. There is a metasearch initiative going on right now in the National Information Standards Organization (NISO), which is trying to respond to libraries' request to have a small number of services for proprietary content, rather than a long list of 60 providers. Reed Elsevier is working together with NISO and other large publishers to develop open standards, so that any metasearch company could come in and search these proprietary services.

Finally, Reed Elsevier offers a service called Scirus (www.scirus.com), which does provide access to some of that hidden content that Google does not provide access to. It has scientific papers from the Web, with 150 million Web documents. It has all the Reed Elsevier proprietary content, plus whatever it can license from other publishers.

Most of the proprietary content is available for a fee. The abstracts are free, but if you click through to that content you go right through to the full text, if your site is licensed, and if not, there is a pay-per-view model, at least for the Reed Elsevier material. What Google has in fact done is to create a successful application for open content; however, something equivalent is not considered desirable, to some extent, on the licensed-content side.

Improving Assessments of Publications

Donald King noted that there is another dimension that needs to be considered in assessing the materials that are published. That is the dimension of time following the publication of the articles, or the point of their availability in the preprint archives.

The median age of a citation is approximately six to eight years, depending on the field of science. Most of the reading, about 60 percent, that takes place is within the first year of publication, but almost all of that is for the purpose of keeping up with the literature and knowing what peers are doing. As the age of the material gets older, the usefulness and the value of that material ages as well, but about 10 percent of the articles that are read are over 15 years old. That is particularly useful information in industry, where people are assigned a new area of work that requires them to go back into the literature. Such feedback on uses can help make the publications better.

The Issue of Cultural Diversity in Publishing

Michael Smolens, with a company called 3BillionBooks, said that up to this point in the symposium he had not heard a term that might be referred to as cultural diversity. There are a lot of different cultures and language groups in the world that have a lot to say about many of the issues being discussed at this meeting.

There was an organization founded in 1998 called the International Network of Cultural Diversity. It was started by someone in Sweden who got 30,000 artists, writers, and musicians together because they could not deal with the European Union in their own language of Swedish. This posed a problem for a broad range of interactions on many issues. Their goal is to try to have the subject of very small cultures and language groups be heard at international meetings and consortia, so that when the World Trade Organization is dealing with trade issues, someone there is at least thinking about the fact that language groupings are disappearing very rapidly and cultural diversity should be maintained. The cultural diversity issue around the world is a very sensitive one that everyone needs to keep in mind when discussing publishing on the Internet.

7. WHAT CONSTITUTES A PUBLICATION IN THE DIGITAL ENVIRONMENT?

INTRODUCTORY REMARKS

Clifford Lynch, Coalition for Networked Information

In this session, we start taking a look at the issue of what constitutes a publication in the evolving digital world, with specific attention to the frameworks of science and journals. Although we are focusing on science, there are a lot of developments in scholarship broadly beyond science, technology, and medicine. For example, the humanities have been very active and creative in exploring the use of the new media.[22]

As we look at this question of what constitutes a publication and how the character of publications changes, we switch our focus from the environmental questions of publication as process that the previous panel session discussed to how we author and bind together pieces of authorship into structures like journals.

We can approach this from two kinds of perspectives. One is from the individual author's point of view. The practice of science is changing, as is well documented in the National Science Foundation (NSF) cyberinfrastructure report, for example. It is becoming much more data intensive. Simulations are becoming a more important part of some scientific practices. We are seeing the development of community databases that structure and disseminate knowledge alongside traditional publishing.

From that perspective, one might usefully ask questions about how people author articles, given the new opportunities technology is making possible. It is clear that articles can be much more than just paper versions by digital means. In fact, they are often printed for serious reading and engagement, and the journals still are using all of this technology around an authorship model that is strongly rooted in paper.

There are trivial extensions that could be made, but there are less trivial extensions, too. We can, of course, add multimedia. Moreover, not all of our readers are going to be human. Software programs are going to read the things we write, and are not very bright sometimes, so you have to write differently for them. So this is one perspective, the author's, that we can look at.

The other perspective is that of the journal, of the aggregation of these articles, recognizing that the ecology in which journals exist has changed radically. It used to be that the other items in that ecology were other print journals and secondary abstract and indexing sorts of databases. Now it has become very complicated. There are all kinds of data repositories. There are live linkages among journals. There is an interweaving of data and authored argument that is becoming very complex. These are the kinds of issues that we will have an opportunity to explore in this session.

---

[22] See, e.g., Roundtable on *Computing and the Humanities*, jointly sponsored by the National Research Council, the Coalition for Networked Information, the National Initiative for a Networked Cultural Heritage, and the Two Ravens Institute, which can be found at http://www7.nationalacademies.org/CSTB/pub_humanities.html.

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

Monica Bradford, *Science*

This presentation about *Science*'s Signal Transduction Knowledge Environment (STKE)[23] summarizes its history and current status, the specific issues related to defining what is a publication in the digital environment, and then how *Science* has used its power as a more traditional publication to help move forward this less traditional project.

The project was started in 1997. At the time, the staff at *Science* thought it was a bold experiment. It was formed jointly by the American Association for the Advancement of Science (AAAS), Stanford University Library, and Island Press. The reason the three groups came together was that Stanford University Library was very interested in making sure that the not-for-profit publishers and the smaller publishers were able to be players online as we moved into the digital environment. They also had started up HighWire Press. AAAS had recently launched its electronic version of *Science* and was excited about the possibility of working with Stanford University Press on new technology ideas. Island Press is a small environmental publisher, primarily of books, but they had ties to the Pew Charitable Trust, which was interested in funding some kind of publishing experiment online. Island Press was helping to determine what the right area for that might be. They were particularly interested in the intersection of science and policy. Of course, AAAS was so excited about creating *Science Online* and all the potential that the online environment might offer, that it was eager to try something new.

The goal of the knowledge environment was to move beyond the electronic full-text journal model. The idea was to provide researchers with an online environment that linked all the different kinds of information they use, not just their journals, together so that they could move more easily among them and decrease the time that was required for gathering information, thereby giving them much more time for valuable research and increasing their productivity.

Why was signal transduction the first area that was chosen? The funders of the project wanted to find an area that would have the chance to become self-sustaining. Therefore, science at the intersection with policy was quickly eliminated, particularly because so much of the literature in that area is actually gray literature, not digitized, and had unclear prospects. The project moved instead to an area where AAAS and *Science*, in particular, were very comfortable. Signal transduction is a very interdisciplinary field within the life sciences. Cell biologists, molecular biologists, developmental biologists, neuroscientists, structural biologists, immunologists, and microbiologists all come to a point in their research when they need to know something about signal transduction.

There also were some business reasons, not necessarily cost or revenue, but the kinds of factors for which a publisher typically looks. It seemed there was a broad potential user base, with both industry and academia very interested in this topic. There was no primary journal at the time, with the information spread across a lot of journals, nor was there a major society.

Other aspects about this area of research and the kind of information in it were some of the most important reasons for the partners wanting to pursue it. The area of signal transduction is very complex and the information is widely distributed. It was important to be able to create links between these discrete pieces of information to help push knowledge forward in this area. It appeared there was the potential by making these links for substantial gains in practical and basic understanding of biological regulatory systems. In short, it was an ideal place for AAAS to begin such an experiment, because it would reach across disciplines, and after all, that is what AAAS is all about. The information in signal transduction had outgrown what could be done in print, and it really called out for a new way of communicating.

One thing the STKE partners were somewhat surprised to find out was that not only did they have to answer the question why signal transduction is important, but for business reasons they had to answer what signal transduction is. Although it was very clear to researchers what it meant, in the business world, they had to explain why a library should care about a knowledge environment around this topic— that a lot of their different researchers, schools, and departments would be interested in this. So, there had to be an education effort that went along with the marketing process.

---

[23] For additional information about *Science*'s STKE, see http://stke.sciencemag.org/.

The STKE was the first of the knowledge environments hosted at HighWire Press. There are now five. AAAS considers STKE as part of the suite of products referred to as *Science Online*. One of the areas of knowledge environment that *Science Online* is moving into next is the biology of aging, because it has some of the same characteristics that were used for choosing signal transduction before: It is interdisciplinary, does not have a home base, and a need exists for scientists to talk across fields.

The STKE has both traditional and nontraditional types of publications and functions. In addition to providing access to the typical journal literature, *Science* tried to create a community environment, so there is an area with tools that relate to that community, resources that scientists would use, and then the most interesting part, which is the various external connections.

On the macro level, what is a publication? Parts of STKE are very familiar; they look like what you would think of as a publication, although it is really a collection of many items that used to be considered each as a separate publication. The STKE has connected them all together in this environment, which is considered to be a publication itself. It has its own International Serial Standard Number (ISSN); it is indexed by Medline; and the updated reviews, protocols, and perspectives all have their own digital object identifiers. In many ways, that part of it is very familiar and very similar to a traditional publication, but then it was combined with these other aspects to make a larger publication. The scientific community will see more of this in the future, as publishers move away from just putting the content of print journals online and try to pull diverse sources together.

The STKE virtual journal has full-text access to articles about signal transduction from 45 different journals, including some that are represented at this symposium. They are referred to as the participating publishers. When their journals go online at HighWire Press, the STKE uses an algorithm that scans the content and selects the materials related to signaling, which the subscribers to the STKE can access.

The community-related functions of STKE include letters, discussion forums, and directories. That has been the hardest part to develop. The initial presumption was that this would be perhaps the most exciting or interesting aspect, providing people with a place to talk across fields and to each other and with scientists that they do not normally see at their meetings, but in fact, that has been the most difficult part to really get going. Other STKE functions include learning services, reviews of the current literature with highlights of what is important, and custom electronic table of contents and personalization—the ability to put things in personalized folders, to save searches, and to relate resources.

At the macro level, the STKE has been accepted as a publication. It has 45,000 unique computers visiting a month, with about a quarter of those coming back more than once a month. There are 30,000 requests for full texts of original articles, 5,000 PDF reprint downloads, and 10,000 connection map pages viewed.

Within this larger publication there are parts than in and of themselves are a new kind of publishing and a new kind of publication. One of these is the connections map. This is basically a Sybase relational database of information about the signaling pathways. The connections map is pathway-centric, versus some efforts that are being done that are molecule-centric. In the long run, this will be an interesting way of synthesizing information and will have a lot of use.

Each entry within the database is created by an authority, who is solicited by the editors at STKE. On top of this database there is a graphical user interface. To create this database, HighWire worked with STKE to create a software that is downloaded for the authority to work on to enter data into the system. As time goes on, the real value of this will be adding bioinformatic tools on top of this database to enhance the ability to find new and interesting information. The authorities that work with STKE on this are willing to do so because *Science* had existing relationships with them as authors. They trusted STKE to put out quality content and to try to make sure that this kind of effort is recognized. They put this kind of effort into creating the STKE because of the reputation of *Science* and its reliability. During this time of experimenting, that is really important.

At this point, the question is: Is this effort moving beyond what would just be called a database entry to a true publication? For each of these pathways, the authors are supplying abstracts, and metadata are being created at the pathway level. The STKE plans to submit those to PubMed and see what happens when it submits the rest of its information.

Each pathway has an identified author, who is an authority, and who can work with a group. Other people may be brought in because some of these pathways are so complex that more than one authority is needed. They synthesize and vet vast and sometimes conflicting literature. The STKE edits the data in the database before it is released and approved, though this process is far from perfect, and it is more difficult than anticipated.

These pathways are reviewed each year just before *Science* publishes its special issue that focuses on the pathways. The review process is very hard, because one has to go through a lot of information. It is all networked. One of the things that the STKE needs to develop is tools for the reviewers—how to be able to know systematically what they looked at, how it is connected, at what point they saw it, and have some sort of printed output for them to look at, because navigating through the database is difficult. At the bottom of each information component, there is a date and time that it was updated. Then, the "Viewpoint" in *Science* provides a snapshot of the state of the knowledge in this pathway at the time of review. The pathways can be reviewed continually, however, if there are significant changes to the pathway or changes in the authorities that are involved.

In summary, the STKE is trying to create new knowledge and look for ways to explore the network property, the signaling systems that cannot be done in the print product. One needs to look for interpathway connections, to clear the pathways and look for networking properties, and then use this tool for modeling and integration with other bioinformatic resources.

There are still many inputs that need work. The STKE project plans to look for more funding to do some of these things. The goal is to make it easier for scientists to discover new knowledge and to be more productive.

There have been some lessons learned in the STKE experiment already. The definition of a publication is evolving. The researchers understand how this information can work for them, and they are excited. That is the best part of this and what makes it fun, because they say what they need.

Efforts at standardizing data input and control vocabularies have been difficult. NIH tried to help with this, but the standards are not evolving. Someone will have to just start doing it, and from there the standards will start to take hold.

The reward system is not yet in place for those people who are doing this kind of authoring. For that reason, the STKE felt that it had to link this work to the traditional mode of publication during the initial transition phase, so that the contributors get credit.

Finally, the "Viewpoints" in *Science* do get cited. They are in PubMed and they draw attention to the site. Hopefully, over time, we will see the pathways themselves cited more than just by *Science.*

PUBLISHING LARGE DATA SETS IN ASTRONOMY—THE VIRTUAL OBSERVATORY

Alex Szalay, The Johns Hopkins University

Why is the publishing of large data sets an issue? Scientific data are increasing exponentially, not just in astronomy, but in science generally. Astronomers currently have a few hundred terabytes derived from observations made in multiple wavelengths. Very soon, however, they will start projects that will reobserve the sky every four nights, to look for variable objects in the temporal dimension. At that point, the data will increase to a few petabytes per year.

Astronomers are searching for new kinds of objects, for interesting ones that are atypical. They are also studying the properties of the typical objects, and for every object they detect in the sky they derive 400 different attributes.

The bottom line is, the volume of data doubles every one to one-and-a-half years. Astronomy, as most other fields of science, operates under a flat budget, so astronomers spend as much money as they get from their funding sources to build new observational tools and computer equipment to get more data.

The data in astronomy typically become public after about one year, because there is an initial proprietary period of use for the people who build the observing instruments and who schedule observing time, but after that the data are released. Everybody can see pretty much the same data that are publicly available.

How will astronomers deal with this? The transfer of a terabyte of data at current Internet speeds takes about two days, but when we get to a petabyte, it will take years to do anything with the data in the

networked environment. This is already on the horizon, and, by the way, a petabyte currently would be on 10,000 disks. Consequently, astronomers need to create databases that are searchable so they do not have to look at all the data at once, and can use indices to find what they need in the data. They also can do searches in parallel that way.

The driving force for astronomers is to make discoveries. They are trying to go deeper in the universe and also trying to explore a broader range of wavelengths. There is an interesting parallel to Metcalf's law. Robert Metcalf worked at Xerox PARC and is one of the inventors of the Ethernet. He postulated that the utility of computer networks grows not as the number of nodes on the network, but as the number of possible connections we can make. The same thing is true with data sets. If we have N different archives, we can make order of N squared connections between those different data sets, so there are more things to discover. That is the utility of federating data sets.

The current sky surveys have really proven this. Astronomers have discovered many new objects, where they combine multiple surveys in different colors. There also is an increasing reuse of scientific data, so people are using each other's data for purposes that they were not necessarily originally intended.

The data publishing in this exponential world is also changing very dramatically. All these big astronomy projects are undertaken by collaborations of 60 to 100 people, who work for 5 or 6 years to build an instrument that collects the data, and they then operate it for at least that long, because otherwise it would not be worth investing that much of their time to do it. Once they have the data, they keep using them and eventually publish them. They put the data in a database and make them accessible on their Web sites. When the project ends, the people go on to other projects, and at that point they are ready to hand over the data to some big national archive or centralized storage facility. After that, the scientists are still interacting with all the data.

Why are the roles changing? The exponential growth makes a fundamental difference. First of all, because these projects last six years, or more, at any one time the data at the national data facilities are only going to hold about 12 percent of the data that have been collected, and everything else still remains with the groups of scientists. This is very different from the previous linear progression of data collection.

There is also more responsibility placed on the projects. The astronomers and other scientists are learning how to become publishers and curators, because they do not have a choice if they want to make their data public. This is Professor Szalay's situation. He was trained as a particle physicist, who turned into a theoretical cosmologist and then became an observational cosmologist, because that was where the new breakthroughs were occurring. Now he worries about data publishing and curation, because this is necessary to do the science. He spends an increasing fraction of his budget on software, and in many cases reinventing the wheel. More standards and more templates would help with this.

Luckily there are many emerging concepts and developments that help. One is that it is becoming easier to standardize distributed data, for example, using XML. There are Web services emerging, supported on many platforms, that make it very easy to exchange data, even complex data.

There also is a major trend in making computing more distributed. This is called grid computing, where the computing is distributed all across the Internet on multiple sites, and people can borrow time on central processing units (CPUs) whenever they need it. The people who talk about grid computing, however, tend to think only about harvesting the CPUs; they do not think about the hundreds of terabytes or possibly petabytes of data behind it, because we currently lack the bandwidth and cannot move the data to the computers. If you need huge amounts of data where every byte needs a little bit of computing, it is not efficient to move the data to the computer site. It is more efficient to move the computers and the analysis to where the data are.

Essentially, we now have an intercommunication mechanism, and a lot of what is done in grid computing also applies in this distributed work, which is growing in an exponential way. It is also getting exponentially more complex. The threshold for starting a new project is getting lower and lower as the threshold for the hardware is getting cheaper.

Professor Szalay got into this through the Sloan Digital Sky Survey, which is sometimes called the Cosmic Genome project. It is one of the first big astronomy projects that is set up in that mode, to map the sky not just to do one thing, but to try to create the ultimate map. There are two surveys being done. One is taking images in five colors and the other is trying to measure distances. There is quite a lot of software involved. At the time when this project started, which was in 1992, 40 terabytes of raw data looked like an enormous amount. Today it does not seem like such a big deal. In one night the

84

survey takes a 24,000 by 1 million pixel image of the sky in five colors. When it is finished in 2005, there will be approximately 2.5 terapixels of data.

The project is also trying to measure distances so that through the expansion of the universe, astronomers can figure out the distance of the galaxy. For about 1 percent of the objects they are trying to get very detailed information.

The data flow is such that they take the data on the mountaintop in New Mexico and ship the tapes via Federal Express to Fermilab in Batavia, Illinois, because the networks are simply too slow to move that much data around. Then they process the data and put them into an SQL database, which currently has about 100 million objects in it, and will have about 300 million when the project is completed.

Professor Szalay and Jim Gray, of Microsoft Research, have begun a project to make these complex data understandable and useable by high school students. They opened a Web site in 2001, and after two years they had about 12 million page hits, and now get about 1 million page hits per month. It is used by high school students who are learning astronomy and the process of scientific discovery, using up-to-date data—data that are as good as any astronomer can get today—and also by professional astronomers, who like the ease of use.

The project identified other issues that they are just starting to appreciate. After they released the first set of data, which was only about 100 gigabytes, they realized that once they put out the data, people started writing scientific papers about them. They are putting out the next data release in the summer of 2003, close to a terabyte, but they still cannot throw away the old data set, because there are papers based on it. Someone may want to go back and verify the papers, so whatever they put out once is like a publication; they cannot take it back. They have a yearly release of the data on the Web site, which is like a different edition of a book, except that the book is doubling every year.

The project also brings up other interesting aspects. Once the project goes offline, the databases will be the only legacy. Most of the technical information is going on in e-mail communication. So they also have to capture, for example, all the e-mail archives of the project. They should not delete them, but rather add them to the database, because this will be the only way that somebody years later can figure out what they did with some subject, or a certain part of the sky.

As Jim Gray says, astronomy data are special because they are entirely worthless. He means this in a complimentary and good sense. He works for Microsoft, so he does not have to sign disclosure agreements and bring in lawyers if he actually wants to play with some astronomy data. There are no proprietary or privacy concerns. You can take the data and give them to somebody else. They are great for experimenting with algorithms. They are real and well documented, spatially and temporally dimensional. One can do all sorts of exercises with them that one has to do with commercial data, but without being sued. Astronomical observations are diverse and distributed, with many different instruments constantly observing the sky from all the continents, in different wavelengths. The questions are interesting, and there are a lot of data.

This all adds up to the concept of a "virtual observatory." Szalay and Gray were struggling with their data publication process, based on a relatively small survey. Their colleagues at the California Institute of Technology, the Space Telescope Science Institute, the NASA Goddard Space Flight Center, and various places were all doing the same thing. They all decided that it is much better to try to do it together because, eventually, astronomers would ask why it does not work together.

When they created the concept of the virtual observatory, the vision was to make the data integration easy by creating some standard interfaces and to federate the diverse databases without having to rebuild everything from scratch. They also wanted to provide templates for others, for the next generation of sky surveys, so astronomers could build it the right way from the beginning.

This idea has taken off. About a year and a half ago, NSF funded a project for building the framework for the national virtual observatory, which involves all the major astronomy data resources in the United States—astronomy data centers, national observatories, supercomputer centers, universities, and people from various disciplines, including statistics and computer science. This project has already developed some demos, which led to some unexpected discoveries. It is also now growing internationally. This effort is now being copied in more than 10 countries, including Japan, Germany, Italy, France, the United Kingdom, and Canada.

Today, all these projects are operating with a funding of about $60 million, and there is really active cooperation. In late April there was a one-week-long meeting in Cambridge, England, about the standardization efforts—what are the common dictionaries, what are the common data exchange formats,

85

how to do common registry formats that are OAI compatible, and so on. There is now even a formal collaboration—the International Virtual Observatory Alliance.

Publishing this much data requires a new model. It is not clear what this new model is, however, so astronomers are trying to learn. There are multiple challenges in the use of the data for different communities. There are data-mining and visualization challenges—how to visualize such a large, distributed complex set of databases. There are important educational aspects of it; students now have the ability to see the same data as professional astronomers. And there is very much more data coming, petabytes per year by 2010.

Astronomy is largely following the path of particle physics, with about a 15-year time delay. Particle physics also grew through this evolutionary phase. It will probably last for the next 10 or 15 years, until the next telescope will be so expensive that only the whole world together can afford to build one, as happened with the CERN LHC accelerator today. Until then, there will continue to be this data doubling.

Indeed, the same thing is happening in all of science. We are driven by Moore's law, whether high-energy physics, genomics, cancer research, medical imaging, oceanography, or remote sensing.

This also shows that there is a new, emerging kind of science. When you think about how science began, it basically was very empirical. It started with Leonardo da Vinci, who did beautiful drawings of turbulence and described the phenomena as he saw them. Then through Kepler, through Einstein, people wrote down the equations that captured in the abstract sense the theoretical concepts behind the phenomena and the natural world, and provided a simple analytical understanding of the universe around us. Then a computational branch of science emerged over the past 20 or 30 years, and what we are faced with now is data exploration. We are now generating so much data, both in simulations and in real data, that we need both theory and empirical computational tools, and also information management tools to support the progress of science.

GENOMIC DATA CURATION AND INTEGRATION WITH THE LITERATURE

David Lipman, National Institutes of Health/National Center for Biotechnology Information[24]

David Lipman recently met with Jim Gray and Alex Szalay. A stimulating discussion on the similarities and differences between biomedical research and astronomy ensued.

As Alex Szalay noted, one of the driving forces for most scientists, certainly those in biological research, is that science is becoming more data intensive. This means that researchers are generating more data for each paper, but they are also using more data from other scientists in their own research. That has an impact on both the factual databases and the literature. Many scientists believe that in order to make the most of these resources we will need to have deeper links and better integration between the literature and the factual databases. This includes improving data retrieval from both types of database, improving their actual usability, and maximizing the value that can be extracted from them. The quality of the factual data can be very much improved if tighter integration can be achieved between the literature and the databases.

The growth in the number of papers in PubMed and MEDLINE is substantial, but it is basically linear. If we look at a number of factual databases, however, in most areas of biology the amount of data is increasing exponentially. For example, this is true of both the number of sequences and the total number of nucleotides in GenBank.

An example of postgenomic research that generates a lot of data is proteomics. If you look at the number of proteins reported as identified in earlier studies, and compare this to the number reported in more recent articles, the amount of data generated with each paper is increasing. Most universities now have a variety of proteomics core services doing mass spectrometry that produces a lot of data.

In the area of expression analysis—the kind of work that Pat Brown has pioneered—one can see the same kind of pattern. There are now many more labs doing this work, but also, any given lab can generate more data points per paper as the cost of doing this kind of experiment goes down. For example, an expression analysis experiment on the human genome may use a chip that monitors the

---

[24] See http://www.ncbi.nlm.nih.gov/ for additional information about the NCBI.

expression of a few thousand to perhaps 20,000 genes over a range of conditions. Hundreds of thousands of data points could be generated that are relevant not only to support the results reported in a scholarly paper, but it would also be useful to make those data available so that others can learn new things.

Not only are scientists producing more data, but they are also using the data generated by others for their own research. There are more than 400,000 users per weekday at the National Center for Biotechnology Information (NCBI) Web site, and this number is growing; most of these people are using data to design experiments. As well as accessing information though direct searching of databases, papers in electronic journals now have many links to, or at least cite, the identifiers of databases records that the authors used in writing their papers. Supplementary data files are also increasingly submitted with research articles, and these too are frequently linked back to the source information.

A typical functional genomics approach, such as gene expression analysis, generally requires a range of genomic data to set up the experiments - sequence data from a number of transcripts or genomes are needed to design microarrray chips. In proteomics, it is essential to be able to compare mass spectrometry data against a number of genomic data sets in order to interpret them fully.

Another important result of functional genomics processes is that the researcher will often generate a new kind of data. In proteomics, the new data might be on interacting proteins. These data may also be relevant to the very databases that were used to design the experiments. For example, a researcher doing proteomics in a microbe may get mass spectrometry data that confirm that certain proteins are actually translated and found in the cell. That information needs to be transferred back to the relevant databases so that researchers using the databases in the future will know that some previously hypothetical genes are now confirmed, and are translated or expressed.

A related point is that if the researcher who generates these data keeps them only as supplementary data files on a personal server, then they are not going to be structured consistently, and therefore the data are not going to be as useful as they would be in a public database, where data are structured and normalized. For expression analysis experiments, some journals are just beginning to require submission of the data to various archives. It is very important to keep in mind this cycle of feedback between source data and publications, because it will affect the way the literature is presented in the future.

At PubMed Central,[25] which is the National Institutes of Health (NIH) archive for the biomedical literature, there are many links and other functions. For example, it is possible to link from the cited references in a full-text article to the appropriate PubMed records, as well as to a variety of other databases. One can now also look computationally through the text for matches to known organism names, and if the names appear in certain places in the article, for example, a Methods section, this increases confidence that the paper contains relevant information with respect to that organism. By having this fairly fine level of integration and links between the literature and factual databases, the article has a higher value: not only can the reader better understand the point that the author was trying to make, but he or she can also go beyond that point and look at newer information that may be available.

DISCUSSION OF ISSUES

Subscription Model as a Restriction on Use of the STKE

David Lipman began the discussion by noting that because the STKE is proprietary and requires a subscription to access it, people do not link into it as much as they might otherwise. It is a limitation when you have some of these factual resources that not everybody is able to get into.

Monica Bradford responded that the connections database is free and anyone can use it. The other parts of the site are behind a subscription wall, those that are more your typical journal items. One of the ideas behind the STKE was to see if you can support the whole combined environment while keeping the database itself free. That is what all the authorities want.

The data most likely also will have value for drug discovery. One of the ideas is to see if STKE

---

[25] For additional information about PubMed Central, see http://www.pubmedcentral.nih.gov/.

can license the data for use by pharmaceuticals companies so they can integrate them with their own databases on a proprietary basis, but providing enough support for the overall site to allow it to remain free for academic use. So far, most of the STKE's experts, the authorities, are comfortable with that model. It is a little early to say how much traffic STKE will have, but it is expected to grow a lot.

AAAS also has an NSF grant for Biosite Ednet, which is trying to take online resources and make them available, or make it easy for instructors to find them to use them in their course work. The STKE has adopted the same metadata that are being developed for that program. This has allowed STKE to extend its audience from primarily researchers, who were the initial focus, to incredible use in education in undergraduate courses.

The Need for Federal Coordination and Investment in the Cyberinfrastructure

Dan Atkins said that the three excellent presentations in this session were representative of the 80 or so exciting testimonies that his committee had as part of the cyberinfrastructure investigations at NSF. He would like to encourage the scientific community to get behind NSF to do something bold in this area.

First, the exponential data growth is now present in many fields. It illustrates that the challenge and opportunity includes going to higher performance networks, higher speed computers, and greater capacity storage, but to do that together with another dimension that he mentioned earlier, and that is functional completeness, by having the complete range of services. The challenge involves this balance between increased capacity and increased function.

A second point that the preceding talks illustrate is the exciting potential for multiuse technologies—the fact that the same underlying infrastructure is serving the leading edge of science, and making the learning of science more vivid, authentic, and exciting, all at the same time. Although a major investment is needed to create this infrastructure, once it is created, as the astronomy example illustrates, leading-edge teams or individual amateurs can make seminal and important contributions to science provided they are given open access to these data and to the tools.

Finally, both the opportunities and the challenges illustrate the urgency for some leadership in this area. The various research communities are already doing this, and they are scraping together the resources. Cosmologists are becoming data curators, and so on. People are putting extraordinary efforts into it, and that is very commendable. At the same time, if we do not get the right investments or the right synergy between domain scientists and librarians and information specialists, we could end up with suboptimal solutions, solutions that do not scale, and worst of all, we can end up with balkanized environments that cannot interoperate and result in enormous opportunity costs. So this is a plea for the NIH, the Department of Energy, and the other agencies to cooperate and try to create the kind of environment that is going to allow these research opportunities to prosper on a grand scale.

Quality Control and Review of Data in Very Large or Complex Databases

Paul Resnick asked about the quality control or review process for the data that gets into the very large or complex databases. Can anyone put their data in?

Alex Szalay responded that typically in astronomy a threshold is emerging. The condition for putting contributed data online is that they are provided with adequate metadata in certain formats. This keeps a lot of people out whose data are not of high enough quality, and who have not documented them sufficiently. This issue was given a lot of consideration, but they did not want to institute a formal refereeing process. They will probably introduce an annotation system eventually, where people can feed in comments and annotations. David Lipman noted that in the databases at NCBI, and typically for molecular biology in the life sciences, there are databases such as Genbank, where authors directly submit sequence data into an archive. When that process of direct office submission started about 20 years ago, the sense was that people were going to put in all kinds of make-believe data and so forth. Actually, that does not happen; although some data are redundant, and some versions are of higher quality than others. That has given rise to related databases, which are curated, some by expert groups on the outside, others by the database managers themselves. So, for example, NCBI curates the human sequences along with some others, such as mouse sequences. There is a comprehensive set of curated

88

sequences.

For the molecular biology data, there are two related sets, the archived set that represents what scientists provided at the time and the curated set that contains what is supposed to be the best version of the data. Those two data sets work together well, because sometimes what some experts think is correct in the curated set turns out to be incorrect and something that was deposited originally may be more correct, so there are pointers to the older versions.

Monica Bradford noted that the STKE is more of a curated database. The pathway and all the related components are externally reviewed by people outside of the group of authorities that created it and did the data entry. That is a snapshot in time, however, and those pathways and data entries need to be constantly updated. Right now, the only way that someone can comment during the period between the formal reviews is either by directly sending an e-mail to the authority, which one can do right from the graphical interface, or by using the feedback function on the Web site, in which case everyone sees it. The STKE will provide some additional tools to improve the review process and also to allow for more community annotation over time, as long as it remains clear which portion is the community annotation and which is the official part that has gone through the review process. A combination of these two approaches will add the most value over time.

David Lipman raised one other point that relates to the difference between astronomy and biology. Astronomy has an organizing principle, in that there are space-time coordinates, which is largely agreed upon. Although there are a number of different coordinate systems that are used, they can be resolved largely to one. Given that there is a stronger theoretical base in astronomy from physics and there is this natural organizing principle, it is possible with a variety of computations to actually assess some aspects of how good the data are.

In biology, at the level of the genome, transcripts, proteins, and, to some extent, protein structure, there are organizing principles that are natural and strong enough to allow detection of data anomalies that do not make sense. The information in a database fits together in a certain way, and as one gets more transcript information or more comparative data for proteins, one can see better how it fits together. Above the level of the genotype and perhaps protein structure, however, with expression data, pathways, and proteomics there are not such natural organizing principles.

One of the difficulties of a project like STKE and other functional genomics projects is that it is much more difficult to use cross-validation to fully assess the quality of the data. Because there are high-throughput methods in biology that are at the level of function, it is a challenge to deal adequately with quality.

Monica Bradford added that one of the reasons the STKE took the approach it did is because all the tools are not yet available. People are trying to make connections across different disciplines. A cancer specialist may find an oncogene that turns out to be in a signaling pathway. They may be looking for information in literature or data that they have not followed before.

The value of having scientists, university libraries, and the STKE editorial processes associated with the curation is that it helps build some trust in the information when the more automated tools that David Lipman mentioned do not exist. You can also link to other archival and curated databases elsewhere, but at a certain point, when you are pushing the edges and trying to gain new knowledge, you have to figure out what you are going to trust. That is what the STKE hopes to improve.

Professor Szalay also observed that as the data grow exponentially, over a period of time they will grow by a factor of 100. The capacity of computers will also grow by a factor of 100, so if the organizing principle is to save every bit of data collected, one can keep up. If the problem is to connect every bit of data to every other bit of data which is N squared, there will be 100 times more data and the computers will be 100 times faster, but the computational problem will be 10,000 times larger. We are starting to approach this.

Martin Blume said he was struck by the connection between a number of things that have been discussed here and things that have happened in the past. It is useful to look at how we got here, and perhaps to extrapolate into the future. The e-Print arXiv grew out of the preprint archives in print back in the 1960s that followed the Xerox revolution at that time. Looking back particularly from astronomy to particle physics, there were experiments done in the 1960s with bubble chambers, and there were many photographs taken.

Those data represented the equivalent of the current sky survey, because they could be used not just for one discovery, but for meta-experiments that were done on that. In fact, one can see the same thing now happening with the astronomical data, where one can reuse them for many different

89

experiments or analyses that yield new data into the future.

The particle physics data were available freely after a period of proprietary use by the investigators (as are the astronomy data today) to others who could use them for new ideas and new experiments. Of course, the digital technologies have changed this dramatically, but it still looked like a massive job in the 1960s. There was one paper where the omega minus was discovered, and 33 people had collaborated on it. It was a sign of things to come.

Data-Mining Restrictions from Proprietary STM Information

Linda Rosenstein, at the University of Pennsylvania, asked how we might automatically download, create, and/or centralize a repository of identified articles if the intent is to extract data and republish subsequently extracted facts. As the extracted data would in no way resemble the original text and would also have pointers to their sources, this should not pose a significant infringement issue. The University of Pennsylvania scientists want to do something that is scientifically quite new, and they think it is very important for cancer research. They know there are various restrictions on the access to the data and the use of the data that their library has licensed. How can they undertake this incredible data-mining process, which presumably will have great results in science and perhaps even in medicine, when they are still subject to the proprietary model of how the scientific literature is made available to them?

David Lipman said that is a good question, whether the current model of fee for access fits how the literature and the factual databases need to be used. However, if the activity in question involves text-mining work based on looking at papers, that will not be useful. If, instead, the researchers want to mine the organized data sets associated with the literature, and would use the two together, that is an issue. It is one of the issues with which the publishing community will have to contend. There is clearly a tension between the current model and how scientists want to work, but it is not clear where it is going to go.

PubMed Central is a service in which the publishers volunteer to participate. They are not necessarily providing the newest versions of their information into it, but scientists request subsets or entire sets of PubMed Central data to compute on locally and to search for discoveries. Some of the publishers that participate in PubMed Central, upon request, allow the download of this information. As PubMed Central begins to get more open-access journals, it will be able to provide the material for download automatically, just as it does with the sequence data. The more literature that is provided under open access the more that allows for new directions in terms of how the literature and the data sets are used, and how they are developed in the first place.

PubMed Central recently interacted with some radiologists who were interested in creating a special database of radiological images, which would be useful for education and training. Clearly, if the journals were open access, then it would be a very natural thing to put the two together. This is a challenge that both the publishers and the scientific community are going to have to face.

Monica Bradford said that this tension is good. It helps publishers to think about what they are doing and the basic goals they have. The purpose of STKE is to help researchers be more efficient. It is supposed to advance science and serve society, and STKE has the potential to help do that. The tension makes publishers rethink their model and experiment with new approaches. It is an evolution. Hopefully, it will be possible to come up with some creative ways that will work in the marketplace, and not be totally dependent on government control or government funds.

Publishing Large and Complex Data Sets for Knowledge Discovery

Participant Eileen Collins pointed out that, presumably, the methods for organizing and labeling the huge data sets reflect current knowledge in the field. Does the availability of all these wonderful data to so many people enable researchers to make quantum leaps in their knowledge of phenomena? Or, is there a risk that because the data are organized according to what we understand now, it might be tempting just to stay in that vineyard and not make big advances or face big changes in modes of thinking? How might this issue be addressed, particularly for disciplines that are not as far along as those that are putting together these huge data sets?

Alex Szalay responded that in the Sloan Digital Sky Survey, which is now about 40 percent ready,

90

after two nights of operation astronomers found 6 out of 10 of the most distant quasars in the universe. This is a small telescope by astronomy standards, but it has proven to be efficient in finding new objects by organizing the data and scanning through the outliers. Is it the only way to actually store the data? The answer is, clearly not. With these large data sets, the only way to make sure the data are safe is to store them at multiple sites around the world. If the mirror sites are organized slightly differently, each enabling certain types of functions, and the queries are redirected to the most appropriate place, it might improve the capabilities.

David Lipman said that for databases like Genbank and for most of the factual databases in the life sciences, there are multiple sites. Genbank, the U.S. human genome sequence database, collaborates with similar centers in Japan and the United Kingdom. The data are exchanged every night, but they have different retrieval systems and different ways of organizing access into them. Furthermore, people can download large subsets or the entire database, and do whatever they want, including making commercial products.

With the STM literature, if there were more open-access journals and multiple sites that provided comprehensive access, one would see different capabilities. PubMed Central is working with a group in France to set up a mirror archive that would organize those data in a different retrieval system. Open access to the data allows for multiple archives that are comprehensive and provides different ways to get at that information.

Dr. Collins's question raised a deeper issue, however. In biology, gene ontology is a way to make it easier for people to see what is there, and to move the information around and to understand it. This represents a trade-off between what we understand now and the kind of new discoveries that Alex Szalay refers to, which change that. There is a tension between reductionism and computing from the bottom up and learning new things, and being able to say this is what we know right now, and having that superimposed from the top down.

Right now, there is a huge amount of interest in ontologies in biology, and some of it may be misplaced. One of the reasons researchers focused on molecular biology was that they really did not understand enough from the top down. If you look at proteins or genes that are involved in cancer, you find those that are a part of glycolytic pathways, and so forth. It is not clear how much these ontologies assist in finding and understanding things, and how much they obscure new discoveries. As long as we maintain access to the data in a very open way, and people can download and do almost whatever they want with them, then if they want to ignore something like an ontology, they can do that.

Monica Bradford agreed with David Lipman, but also made a few observations that are not quite as global, but based on the experience with STKE. The research in signaling started out looking at things very literally. A researcher followed a signal down one pathway, and soon realized these were really networks; one could not think about one pathway, but had to think about all these pathways together and how they intersected and affected each other. This could not be done in print. So the ability to build a database and be able to represent these and then query across the pathways was very useful.

AAAS would like to develop bioinformatic tools that will actually let researchers look for inter-path connections and find new knowledge. Ms. Bradford hopes that eventually these tools will allow researchers to play with the STKE and add in their own information to see if it changes the effect or has an effect on the pathways. The information is vast, and tools are needed to help organize it. Once the organization of the information is taken care of, it gives researchers a chance to begin to think about new things and to look at the information differently. This process can be found in many other disciplines.

Clifford Lynch noted that another striking example of that is the migration away from surrogates to full texts that allow computation. That capability is having a radical effect in many fields. It is useful to be able to find things by doing computation and searching on the full text of scholarship, as opposed to being locked into some kind of a classification structure for subject description that some particular group used in creating surrogates. That is a theme he hears from scholars in every field from history all the way through biology.

Transformation of Archiving in the Knowledge Discovery Processes

Marcus Banks, of the National Library of Medicine, asked a question that he hoped connects sessions four and five. If we are moving from publication as product to publication as process, should we make a similar transformation in archiving? Or should we still want to archive products that maybe are

extracted from that process? An example in STKE would be the "Viewpoints" from *Science*, which are snapshots in time. It seems that archiving the best information in a separate subset would be useful, but then that might be paper bound and published the old way we have done things.

Monica Bradford said that AAAS has been talking about that. Because STKE is constantly changing and the database is updated on a regular basis, they assume as it scales and grows it will increase a lot. Should the STKE Viewpoints be archived on a certain schedule, perhaps every quarter? AAAS is not sure yet what would be the right frequency with which to do that, but the Viewpoints do not have to be paper bound. Right now STKE is in a transition stage, and the scientific authorities still want to get credit and want to be recognized for the effort, because it takes a lot of effort to do these functions well. So the purpose for the Viewpoints is not so much archival, although it does serve that purpose. It is more to give the authorities some recognition.

Dan Atkins noted that Rick Luce touched upon the archiving issue during his presentation. There is an enormous stream of digital objects that could be created by knowledge discovery processes that are mediated through technology. One needs to be able to archive not only these objects, but the relationships between the objects and the social context in what they are created.

We need to start thinking about archiving that includes these temporal streams that come through. One of the most profound ideas about all of this comes from John Seeley Brown, former Chief Scientist at Xerox and Chief Innovation Officer for 12 Entrepreneur, Inc., who says that perhaps the most important aspect of this technology-mediated way of work is not just relaxing constraints of distance and time, enhancing access, and so forth, but actually comes from the possibility of archiving the process, not just sampling the artifacts along the way.

The idea in areas of ubiquitous computing is that people could then come back and actually mine processes and extract new knowledge that otherwise has been left on the table. It is an extension of this whole notion of data mining into knowledge process mining, so it gets very abstract. We can start to see that it is not just fanciful, and it is something to think about. People who are interested in long-term preservation need to consider huge repositories that take into account not only the basic ingredients, but the social processes by which these ingredients are encountered.

Increasing Data and Lost Opportunities

Donald King noted that over a period of 15 years, scientists have increased the amount of time that they devote to their work, by about 200 hours a year. The point is that scientists are now approaching the individual capacity with regard to how much time they can devote to their work. Most of that additional 200 hours per year is devoted to communicating. The number of scientists increases about 3 percent a year, which means that the total number of scientists doubles about every 15 to 20 years or so, but the point was made earlier in the symposium that some of the information we are gathering doubles every year.

It seems that one of the concerns is the limitation of the capacity of the human intellect to work with these data. The National Academies, NSF, and others could focus on trying to increase the number of scientists who work with these data and the infrastructure that can help them work with these data, to begin doubling the scientists every 5 to 10 years instead of every 15 to 20 years.

In response, David Lipman agreed with the basic point Donald King made, but noted that scientists do adapt to dealing with large data sets. If presented with the computing power and the data, scientists ask different kinds of questions. It takes a long time before more scientists within the community shift and start to think of different kinds of questions. A few pioneers start to think a new way, and then it starts to happen. He has a lot of confidence that the more data we generate, if the computers and access to the data are there, people will come up with ways to ask the questions.

Donald King clarified that he wanted to know if there are lost opportunities. It seems as though there must be. David Lipman said that a lost opportunity does exist. NIH a few years ago set up an initiative called the Biomedical Information Science and Technology Initiative (BISTI) to try to get more people involved in computational biology. They came up with a lot of recommendations. However, he thought the goal of BISTI would be to recommend more training for scientists and more money for computational research for the kind of work that Alex Szalay referred to, where one analyzes other peoples' data sets, because there is a huge number of discoveries to be made. There is to some extent a lost opportunity, because there are not enough biologists researching and writing papers that get

published in biological journals. BISTI should have focused more on training people to do that kind of work and on more grants for doing that sort of research.

Professor Szalay also noted that because of this data avalanche or data revolution, if the data get properly published, it will cause another fundamental sociological change. Today in science, many people train a lifetime to build up their own suite of tools that they then apply to the raw data that they can get. Through these digital resources, one can get reasonably polished data, where one can think more about the discovery itself; a researcher does not have to spend so much time with the mechanics of scrubbing the raw data and converting them into useable form. People will be much less reluctant to cross boundaries, for example, if the data are all available in a ratified and documented form.

The Value of Knowledge Discovery

Mark Doyle, of the APS, said he was amazed when listening to the presentations and comments in this session because they make what he does when he publishes simple papers look trivial in terms of the amount of data and text published. He previously mentioned the two to three orders of magnitude difference between pure dissemination and what a publisher might do in creating an archival XML and doing peer review. There is another two or three orders of magnitude increase in what researchers actually are doing with their time. That makes him hopeful that publishing is really becoming more of a minor cost, compared with the cost of doing research and related activities. He hopes that regular publishing would piggyback on these larger kinds of things, since publishing is important, but not nearly as costly as doing the research or accumulating these much more complex kinds of things. He also asked whether there is a transition to where these things become much more primary than the journal articles that come out of them now.

Monica Bradford noted that the 200 hours of communicating is a huge cost. That is the real publishing, the communicating and getting the idea shared. The amount of time someone puts into creating a product, be it a connections map or whatever, is significant. It is a time away from doing basic research. At STKE, they are happy to hear from the authorities that gives them an added value, in that they have to organize their own understanding and framework in their area. But that is a significant cost one cannot dismiss.

Donald King also asked about the cost for doing the traditional paper. He asked if a researcher or some granting agency would realistically be willing to pay $1500, rather than going through a subscription model. The scale of the costs that are involved in traditional publishing is much greater than that.

David Lipman raised the NIH budget as an example, assuming that the total amount that NIH spent and the number of publications that came out of it is increasing. It is probably $250,000 or $300,000. So the amount for doing the publication in some journal is a smaller part of that. There are issues in terms of economic analyses of open-access publishing versus fee for service on the basis of that. However, Monica Bradford and Mark Doyle were referring to how much time all those other things are factoring into this as well. In a sense, the paper does represent what the scientist did; it is the knowledge part of it. That can only be done so fast. So there is a difference between organizing data sets and making them useable to other people, which is a challenge, and finding and extracting what one thinks about that data set and getting it out there.

The Role of Journals in the Successful Development of Databases in Molecular Biology

Bob Simoni, from Stanford University and the *Journal of Biological Chemistry* (JBC), added that the enormous success of factual databases and our reliance upon them actually are the result of the collaborative effort the journals have made with regard to their "no-wall" requirement that the data be deposited as a prerequisite to publication. One might think that is a natural thing, but it is not. In the area of protein structure, for example, the data that crystallographers gathered were held and not deposited. The pressure from peers to a certain extent, but more importantly from journals, resulted in those data being deposited as a condition of publication, which makes the databases in this area as robust as they are. He then asked David Lipman about the status of a centralized, normalized system for deposition of arrayed data and gene expression data.

93

David Lipman responded that people at the NCBI have been discussing this with David Klein, who is active with the JBC on that issue. There are standards that people have tried to agree upon for what is the minimum amount of data necessary to be submitted to a database. Unfortunately, that minimum is so high that a number of scientists are not willing to do that. Gene Expression Omnibus (GEO), which is an NCBI database, has a somewhat lower threshold in terms of the requirements for submission, but the NCBI is in discussions with the international group on that issue. GEO is growing faster now. Some journals require submission.

He also seconded Bob Simoni's point about the critical role that journals have in terms of getting data in a useful form into these databases. Despite the fact that databases are useful, scientists often do not want to spend their time on data submission. What they are judged by is not what they put into a database, but what they publish. The role of the journals is absolutely critical, and JBC was one of the real pioneer journals in pushing submission to the sequence databases and getting essentially 100 percent compliance.

94

8. SYMPOSIUM WRAP-UP

MODERATOR'S OVERVIEW

Mary Waltham, Publishing Consultant

In the context of Dr. Duderstadt's opening remarks to this symposium, which referred to the "chaos of concerns" that currently surrounds scientific, technical, and medical (STM) publishing, this final session reflects on some key issues raised by the symposium's participants, including publishers, university faculty, administrators, and librarians.

From a publisher's perspective, changes occurring to publishing systems driven by electronic journals have been discussed extensively during the symposium, and a range of issues that require close attention were identified and addressed. It is clear that the costs of online-only journals are less than print plus online. However, print seems unlikely to go away completely across all of STM publishing in the near future, so the costs of a dual system persist for most publishers. Archiving was discussed and questions raised about who will be responsible for the archive journal content—publishers, national libraries, or other third parties? What will be archived—all of the journal content or just the research content? Who will curate the journal archive and ensure it migrates to appropriate platforms as technology evolves? Who will pay?

Within STM journal publishing there are economies of scale that tend to favor publishers of large numbers of journals; these economies of scale are not available to small and mid-sized publishers. As a result, some cooperation and grouping of content have arisen within some sectors in order to mimic or replicate these economies. Recent mergers and acquisitions within the STM journal market have further made smaller publishers nervous about their longer term viability as independent entities. Most of the broad range of online business models are quite experimental and seem likely to diversify and hybridize over time as the publishing system develops and new advances emerge.

Speakers also talked about the likely development of publishing as a more disaggregated process with separate pieces of the continuum done by different groups, from content creation through dissemination.

Filtering and quality control of the information are central to the publishing process but, in the future with more open access to all types of information and data, who will provide reliable and consistent filtration and quality control, and who will pay for it?

Increased online access results in increased usage of information. The journal as a package of information is not granular enough and so further unbundling of information is clearly taking place. Customers and users want more granularity in the online environment than a journal issue represents.

There was discussion of who needs copyright as opposed to "wants it" or "uses it." Copyright is very dependent on both the author and the mission of the publisher with whom they may be working.

The continuous online publishing process means that documents may no longer be static but evolve through time because addition, annotation, and revision are simple.

Interoperability and common standards are essential to bring together and integrate information and provide a dynamic reference tool. Achieving this type of integration is necessary for making the optimal use of scientific information.

A key point for publishers is: Where do they add value to the publishing process, and is that value added where users and customers want it? The role of publishers must continue to change to meet the needs of the research community.

COMMENTS BY PANEL PARTICIPANTS

Malcolm Beasley, Stanford University

Professor Beasley commented on the symposium from the perspective of an interested practicing scientist working in the lab, with a physical science background. He shared an example of the use of print versus online journals in the physics community in Stanford, where the faculty now overwhelmingly use electronic journals. Only the older faculty still use print copies, and it will only be a very short time before online journals are used universally.

Professor Beasley's first impression was that some sort of broad open access to scientific information bases of various kinds, including journals, is inevitable. It is too important not to have them. This is his view as a practicing scientist, and, although they are younger and naive, that is also the view of the students. It will not be simple to get there, and there are a wide variety of vested interests and issues that have to be accounted for. He believes that ultimately scientists will insist on it.

Nonetheless, he agrees with Monica Bradford from *Science* that this will be a tension-filled transition. That tension is not necessarily a bad thing if it is creative. Tension can also be destructive, however, so it will require strong and well-informed leadership to ensure that the tension is creative for science and all who partake in that enterprise.

His second impression is that "publishing" will not go away because it provides added value. The breakdown of the present system is forcing an examination of what those fundamental values added are in the existing publishing modes. Equally interesting are the alternatives in terms of publishing broadly defined to achieve things such as certification, formatting, and editing.

Most interesting to Professor Beasley are the new modes of value added that will be created. There are some wonderful ideas out there. Maybe three-quarters of them will not work, but it is the quarter that survive and provide the essential value added that will be wonderful to have and very interesting to see evolve.

Also, it is clear that access to information almost universally will continue to profoundly improve in speed and efficiency, and therefore have a similar effect on science. That is basically good, but it is important to deal with all that information and get to it with some sense of the quality, the various other types of value added that might be seen. Hyperspeed in itself is not the ultimate good; rather, it is speeding up the process of achieving understanding, as distinct from the acquisition of facts. We ultimately have to achieve understanding in science, and it may mean different things in different fields. During the symposium, Professor Beasley heard ideas that might implement that, but he did not hear a focus on what might be done to improve that aspect of the scientific enterprise. He thought that aspect of the scientific enterprise should have been brought into sharper focus. People are thinking about it, but it is important to stress that we do need to do that.

Professor Beasley thought that in addition to understanding the information we will be getting, we must find methods to improve and accelerate our ability to form good judgments on the basis of that information. There will be more information, which presumably will be better and received faster. To give a concrete example, he observed over his 30 to 40 years as a university professor that graduate students cannot read the literature. It is too abstract, has too much jargon, or has concepts they do not understand. Maybe this is something more specific to physical sciences, but simple access to it is not going to be enough.

As a university professor, he found some related issues as well. How will the faculty help students deal with the information? They may be more adept in using the tools, but they are not going to be adept in this essential value added. For example, there are personalized searches and virtual journals. There are in *Science* and *Nature* perspectives on some of the articles.   He is an avid reader of those because they help him gain perspective, not only in his own field, but a bit more broadly.

The condensed-matter physics community is considering creating electronic journal clubs. There is a famous journal club in this community where individuals report on a paper selected because of its quality and importance, which is then presented and critiqued. This was a tradition at Bell Labs, which was initiated by Conyers Herring, a truly great physicist, and some of the alumni of Bell Labs are now trying to institutionalize that by putting it online somehow. Professor Beasley does not know whether it will work, but believes it is an excellent thing to try.

96

The knowledge environments and connection maps discussed during this symposium are ideas that illustrate what is possible in this area. As a university professor wearing his teaching hat, Professor Beasley found the implications of electronic publishing and the digital revolution interesting and not topics that have received sufficient attention either in the scientific community or in the universities.

James O'Donnell, Georgetown University

Dr. O'Donnell is one of the longest serving editors and publishers of an open-access online scholarly journal anywhere. His cofounder, Professor Richard Hamilton, and he began publishing *Bryn Mawr Classical Review* (BMCR), an online book review journal for classical studies in the fall of 1990. It has always been free; it has always been on the Internet. They tried selling a paper version for a few years, but that did not go anywhere and ended in 1997. An archival CD was published in the mid-1990s and was a complete flop. BMCR has become the leading book review journal in the classics in the United States and one of the top two or three in the world. The editors have all the worries and anxieties of an open-access publisher, including where the next dollar is coming from, but they are confident they are in business to stay.

We expect, he believes, of a new system that it will drive costs out of the old system. There is a general sense that we pay a lot for information and if we could possibly pay less, we would like to do so. He shared the dreams and imaginings of the early 1990s, when we thought that electronic publication would be somehow magically frictionless and cost free. He realizes that we know it is not, but at the same time we know there are opportunities to reduce some of the costs. The first two sessions of the symposium were particularly instructive in that regard.

O'Donnell treats separately the question of recovering those costs and sharing those costs in an equitable and sustainable manner. The open-access movement, to the extent that it is a movement, tends to shift costs away from the user and toward the producer of information. O'Donnell recognizes the attractiveness of this shift, and has qualms about it. The adventures in the non-open-access domain of journal publishing in the past 15 to 20 years have demonstrated that one of the features of high-priced paper or electronic journals is that they facilitate a form of quality control. When university libraries cut their budgets for serials acquisition and reduce the number of titles they acquire, they are exercising a form of quality control over what they will accept. If the "big deal" for licensing e-journals is fraying around the edges, library resistance will transmit itself back to the publishers as news about which journals are to be sustained and which journals are not. The actual closing of real journals and cutting back of titles by publishers over the past decade has been achieved mainly through the Pacman-like gobbling up of one publisher by another and the recognition that not every title was sustainable in the economic model of the for-profit publisher.

But here O'Donnell identifies the focus of an important disagreement. Some of the symposium's speakers spoke up for the open-access model as delivering superior quality. One can equally argue, he thinks, that the commercial model provides superior quality. As one argues about business models, costs, and recovery of costs, it remains an open question whether a change in the system will improve or degrade the quality of the aggregate system of scholarly communication. As long as that question remains open, he thinks there is no forcing argument to use in favor of one model or the other. A mere economic argument would not be definitive. A good recommendation from this symposium, according to Dr. O'Donnell, would be to continue to assess what the effects on quality of information, timeliness of access, and quality of peer review in the different models would be.

O'Donnell observes wryly that if, as an open-access journal publisher, he is part of the solution, as a provost, he is undoubtedly part of the problem for everyone. Commercial publishers will complain that he does not give enough money to the library, and explain that this is why librarians think the prices of the journals are too much. Open-access enthusiasts will complain that he does not underwrite their experiments in new forms of publication, because he is still paying for the old system. He would like to stop paying for the old system before paying for the new system, and not be caught paying for two different systems at once.

From the viewpoint of a provost, he is struck by the emerging differentiation of the product for which the scholars, scientists, and publishers want him to pay. He was particularly impressed by many of the numbers of costs raised during the first session, especially Mike Keller's findings on the precipitous drop in commercial value of information over the first 12 months of its life. That seems to suggest at least

97

a tri-partition of the kind of information objects that are under discussion at this symposium. At one end, there is the timeliest of information services, providing the linked news from the front as rapidly as possible, with no consideration for archiving or price, but simply getting the news as fast as possible so that science can proceed in its best way. At the other end of the life expectancy of that information, service has become an artifact, something to be preserved, maintained, and sustained long after its commercial life, or perhaps even its scientific life, has been exhausted. Preserving the science that is done today for the historians of science 100 years from now is an important exercise, but the historians of science 100 years from now do not have their budgets yet and cannot come to the table with dollars to pay for that task.

That first information service, O'Donnell believes, tends to be market-based. That last artifact service is not market based at all, but is something done out of *noblesse oblige* for the greater good of the community. Between there is a borderland area, where the information service itself needs to be mediated to those who have limited access to the market. If a researcher has large research grants or is in a major research university, he or she is probably doing well in acquiring information. However, if one is in a developing nation or a small institution or is otherwise disadvantaged, the situation may be very different. He believes that understanding that differentiation of product, which is an increasing differentiation in an electronic environment, increasing among products and increasingly differentiated among disciplines, will be an important part of understanding what a new system of information dissemination can be like.

In the end, O'Donnell thinks himself more tranquil than many other observers. Much of the symposium's discussion has been focused in his view not on imminent fear of negative events, not on reasonable expectation of longer-range problems, but on still shapeless anxiety. As provost, he cannot pay to fend off every piece of unsubstantiated anxiety. Every other department and office of the university comes to him with the same kind of anxiety. Much progress has been made in ten years, he believes, and the discourse has changed dramatically. In that regard, he is optimistic. His question for the librarians, scientists, and colleagues he works with will be, Can you make sure you define what the problem is before you need a solution? He is still not convinced that he can get a good answer to that question.

It is time, he believes, to begin disaggregating the scholarly and scientific publishing crisis, if there is one, into pieces that can be addressed in rational and coherent kinds of ways, as part of the many movements that have begun to emerge in the past ten years, as e-journals have become a reality, as open access has become an astonishing reality, and as the impact on academic culture has been so great. If someone can identify persuasively the problem and say what the characteristics of a successful outcome might be, then they will find him to be a much better provost.

Ann Wolpert, Massachusetts Institute of Technology

The views expressed during this symposium have been largely those of the authors and publishers in the scholarly communications system. Ms. Wolpert commented from the standpoint of the university library, which is responsible for bringing in, paying for, managing, and maintaining access to the kinds of information resources that are the subject of this symposium.

To paraphrase Hal Abelson's earlier comment, the progress of science requires access to raw material, evaluated judgments, and conclusions, to work that is current and previous. This is true in the university, not just in one's own discipline, but in multiple disciplines, because universities are dealing with students who have not yet settled into a discipline or who are working in the interstices between disciplines. So it is the responsibility of university libraries to think across those kinds of boundaries.

It appears that the raw material and the evaluated conclusions coming out of scholarly research are bifurcating into two fairly distinct realms. One realm is a quite tightly controlled, often highly priced peer-review literature, and the other is the minimally controlled, scholar-managed, open-access publication and database regime. Both environments present challenges to universities that are by mission dedicated to providing homes for researchers, educating the students who come to these environments, and managing the information that is created today and yesterday for the benefit of scholars and students, some of whom have not yet been born.

One could generalize from the symposium's discussion that the controlled literature is making every effort to increase that control over the content that they publish, and to expand their reach both in

time and in format. We heard that there is considerable interest on the part of publishers in controlling the archive of those publications, so that they would be responsible for the permanent record of a discipline rather than university libraries. The disciplines or the publishers themselves would thus be responsible over time for the archiving of publications. At least some publishers have also expressed an interest in adding data to that record, which would bring all of these resources under the intellectual property control of the publishers themselves.

This may be a perfectly rational business model from the standpoint of someone who manages a university press and struggles on a day-to-day basis with the finances of publishing. However, it is not a perfectly rational business model if one happens to be on the university library end of this formula and looking at buying these materials from publishers under the kinds of intellectual property controls that are created largely for the benefit of the entertainment industry rather than education and non-profit research.

The university perspective on the value chain of the process by which new information is created is very different from the value chain perspective of authors or publishers. In the value chain from the university perspective, the university builds the infrastructure and provides an opportunity for faculty to develop curricula and conduct research. Universities also attract and admit students, especially the graduate students who conduct some of the research. Universities provide access to information for their communities at a cost that must seem to them to be reasonable, as a percentage of the overall annual operating budget of the university. And they delegate that responsibility to their university libraries, which is where research libraries come in.

Not everyone in research libraries believes that the subscription model as a way of acquiring information is fundamentally broken, although the price structure certainly is. Most do believe that the terms and conditions under which information is allowed into our campuses are substantially flawed, because licensing agreements control, in many cases, how subsets of a community can actually use that information. So it is the combination of the price and the constraints on the use of information that represent problems in the system for us.

Looking down this value chain of universities to the point where faculty are evaluated for publishing their work, one can see that the publication function has been outsourced to publishers for a number of valid reasons. As long as publishers handled the peer review and the publication of the work, and that work came back into the university for the next generation of education and research at a price that was reasonable as a percentage of the annual operating budget of the university, the system worked to everyone's advantage.

Over the past 10 or 15 years the outsourcing enterprise has developed its own independent business model, which does not think of itself as having a symbiotic relationship with the university as much as it thinks of itself as a stand-alone, independent enterprise, able to create and craft its own business future, separate and distinct from what the university wants. It is in that environment that the costs have gone up and that the terms and conditions of use have changed, because there is now a very clear boundary between universities and the groups that capture the work that comes out of universities, publish it, and evaluate the quality of that work.

One of the consequences of this separate and distinct business model, that is now at more than arm's-length from universities, is that some of those businesses doing the publishing and the peer reviewing have come to think of universities as patrons—that is to say, organizations that "owe" the publisher a certain amount of money to support the work. Others think of universities as pigeons—that can be plucked endlessly to support a profit margin they may or may not wish to support if they were asked. A third alternative one hears, is that universities are pirates, stealing information, trying to get for free information that was of high value and added to by the publishers. None of these scenarios makes for a particularly useful set of expectations around which to have conversations, although we keep trying.

The conundrum for universities has several dimensions. For many publications, the costs are simply too high for the value received, and the licensing conditions are problematic, in terms of what we can do, particularly with digital information when it is delivered to our campuses.

As described in session three, the intellectual property environment is not only incomprehensible to the average faculty member and student, but it may place universities at risk. Any legal regime that varies so considerably by circumstance and format is not a particularly useful legal regime for people to operate under. On the other hand, people are afraid to seek definition through litigation; because they do not know what the outcome is going to be. Now there are also new forms of data and information that clamor for attention for curation and funding on our campuses.

Ms. Wolpert shared some concluding observations from the perspective of the university librarian.

99

It is apparent scholarly communication and publication are diverging. At one time, communications with colleagues was through publications; now it is clear that one can communicate outside of the formal publication record. The formal publication record is moving in many cases off to one side, which again affects the question of how much value do we put in the formal record of advances in a discipline, if in fact most of the communication is happening some other way. That is in a sense what MIT's DSpace is about—to capture the communication, not the formal publication record.

We hardly have a clue about what reasonable standards and norms might be for the cost of peer-reviewed publication. Those costs vary tremendously, and we do not know what drives them. We also do not know what drives the cost of print as opposed to the electronic publication. So it is very hard for us to think logically without the kinds of norms and standards that one can get from most other industries.

It is clear that intellectual property law that meets the needs of the entertainment industry and the international publishing conglomerates is not particularly conducive to what goes on in the academy. We do not know what new models of peer review and recognition might be developed for open-source publication, which is an area of real attention.

Finally, Ms. Wolpert took away from this symposium a new, tongue in cheek business model for university libraries—the offshore library model. She will advise her colleagues in the state of Maine, who are dying under the current cost structure for peer-reviewed literature and have had to cancel subscriptions, that all they need to do is go to one of the 130 developing countries where this information is delivered for free and open a branch library; they can have an offshore library.

DISCUSSION OF ISSUES

Dr. O'Donnell added that he was struck by the vividness of the presentations. Ann Wolpert's comments are quite relevant to the divergence of views and the building of collaborative enterprises. That said, he thought the speakers addressed fewer of the problems that might emerge as, for example, those huge databases age. In another ten years, 5 or 10 petabytes will be easy to carry around, but it will still take some housekeeping and maintenance, and will begin to look a lot like what publishers and libraries do.

His question remains, what is it going to take to do the good science and to make sure the good science gets done? That should probably be the bottom-line question and set of priorities.

Universities Role in Data Curation and Management

Ann Wolpert asked Malcolm Beasley as a scientist in a research university what role he sees for the university, if any, in the curation and management of data that emerge from science. She asked this in light of Dan Atkins's comments on the National Science Foundation's cyberinfrastructure report and the question about where responsibility resides for the long-term storage and curation of data.

Professor Beasley thought that question was more appropriate for Mike Keller. He did, however, add that faculty in general do not understand the costs of providing these library services in this modern sense and therefore are not necessarily informed people to make judgments about how these things ought to be paid for and what the real trade-offs are.

Ann Wolpert asked the question in part because she was interested in knowing how and where the discussion can go forward on university campuses.

Professor Beasley, reflecting on his own experience, said that the faculty do need to understand these things better, because if they do not they will go and beat up on the provost. To some degree, it simply is not fair, and to another degree, it must be exasperating to be ill informed. There are a number of areas where the provost will get beaten up, but the point is that with the revolution in the way science is done that it is too important to leave to the provost alone.

Jim O'Donnell thought that the real question was not what the provost wants to do in the abstract, but an empirical question: Where is information best cared for? He did not think we know the answer to that. As a provost, he could understand an argument that would say only the scientists know what good-quality curating looks like and what maintains information in good enough form to be useable, therefore it must be done in-house. He also could accept an argument that the scientists do not have a clue how to

do it technically, managerially, with appropriate metadata, or with appropriate access or preservation, and therefore the curating needs to be outsourced. The question then would be, does one outsource it inside the not-for-profit community, the university, or the larger not-for-profit community, or externally to the market?

Donald King commented that the scientific literature should be made available to the faculty, not just per title, but rather per article. For some journals the cost per article has actually decreased because the size of the journal has increased much more rapidly than the price. Faculty should also be presented with the cost based on the price by use. The number of individual subscriptions by scientists has dropped from 5.8 to 2.2 over the past 25 years. As a consequence, all of that reading has now gone into the library. Scientists have roughly doubled the reading they do in academic libraries, and in special libraries it has increased four or five times. Faculty need the right indicators in a way that is realistic for them to understand.

Mike Keller tried to respond to Ann Wolpert's initial question about the role of the university in data curation and management. He noted that the underwriters who work for Stanford and who want the university to spend a lot of money on insurance say that the collections that have been amassed there in the libraries, now amounting to about 8 million books and probably 30 or so miles of archives and so forth, are valued at $1.2 billion. It is one of the largest assets the university has, other than its land.

In the decade since 1993, when the World Wide Web became massively and widely available, Stanford has spent $73 million on capital projects, the most recent of which was to buy land to construct the first of what might be six or eight modules to store books and other physical artifacts, as well as perhaps to house a very large digital store off campus. In the same period of time, Stanford spent about $5 million, earned almost entirely from the publisher clients at HighWire, putting out about a million articles for them. The cost of information technology and e-publishing today is about 5 percent of the whole budget for publishing.

As we work on this problem of nonphysical, virtual artifacts, the solution lies, as it has in the past, in great libraries, which are archives not built by archivists for their own sake. They are built with the connivance and cooperation of the scholars who most directly receive benefit from them. As the scholars in science, technology, and medicine very clearly see the advantages to some of the research possibilities that the big data sets and collections of electronic resources come to offer, then there will be more impetus and more money to provide them with the research materials, which then become artifacts for preservation over a long period of time. Libraries are preparing to do that, but everything that we have is in embryonic phases, without exception.

Martin Blume noted that on behalf of publishers from learned societies, the message, if somewhat ungrammatically, to the professors, librarians, provosts, and scientists is, "We are you, and you are us." The American Physical Society (APS) is run by professors and laboratory scientists, including the president, the vice president, past president, and president-elect. It is operating with scientists, who have been professors. So the values that we have are the same values that are expressed at this symposium. The APS officers understand some of the problems of the publications perhaps a little better, because they have been thrown into it and are therefore forced to think about it. At the same time, they need the understanding of the scientific community and the universities that this is where they are coming from.

APS's publications committee is chaired by an industrial scientist. Its members are university professors and laboratory scientists as well. Paul Ginsparg is a member of that committee, so there is representation from the electronic archives. All of this comes together at this point, so APS cannot be accused of having a very narrow point of view; it is better educated.

Interaction between Scientists and Archivists

Steve Berry asked Jim O'Donnell whether it is possible to overcome the arrogance of the scientists and the humility of the archivists, so that the natural solution will come from these two groups of people learning to talk to each other, to deal with the problems of these enormous databases, and realize that each brings something, but not a complete solution.

Jim O'Donnell thought the critical intellectual capital will be formed in the dialogue between the archivist and the information scientist on the one hand, and the working researcher on the other. He would like the institutions and the not-for-profit sector to retain control of that process in order to hold on to the intellectual capital. Then he would be willing to talk about who should do the dirty work and make

the things actually operate.

He was brought up short once early in the days of BMCR, when an associate provost, who used to be a librarian, asked Dr. O'Donnell if he could do a good job of marketing to nonclassicists, and O'Donnell replied, "not really" since he knew who all their readers were. The associate provost told him that, "the one thing you know in a library is there are lots of other people who have an interest in the stuff you are doing besides you, your six friends, and the folks you think you are doing this for." That really adds an order of magnitude of value that O'Donnell thinks we always need. Getting the dirty work done adds another value, but it is a very different kind.

Tom Arrison from the National Academies was interested in the idea of the dialogue between the scholars and the librarians. An instructive example is the OpenCourseWare project at MIT, where there was a process that led to this fundable vision, implemented with money from private foundations. Sometimes this dialogue and reflection on a campus can lead to a vision that can bring in the resources to implement it. He asked whether in the panelists' individual institutions that kind of dialogue is going on and is adequate. Also, more broadly, he was interested in what needs to go on between institutions and among the scholarly communities to help promote that.

Malcolm Beasley noted that in terms of dialogue, it depends on definition. At Stanford, there is a faculty committee that interacts with the administration to deal with these questions. The faculty hears reports, and he has no doubt that it is a substantive discussion. But a wider discussion than that through the normal faculty senate committees is not widespread. The question is: Is this an issue that is sufficiently important that one ought to try to have a wider discussion about it or not? He would argue that it is a candidate, because of its importance to the scholarly side of what faculty does. But he does not think it would be fair to say that this discussion is highly developed in any institution.

Jim O'Donnell said that the learned societies play a stronger role than the individual campuses do. They can aggregate more resources, and they can bring more intellectual firepower to bear on discipline-specific kinds of questions than happens in a university faculty with a library advisory committee with different disciplinary representatives.

Ann Wolpert believed that perhaps one of the points to take away from the symposium was that there need to be vehicles to encourage that discussion, because there are conversations within institutions and groups of institutions, and there certainly are conversations within disciplines, and, to a lesser degree, among and between disciplines. But there is no easy way for the conversations to happen across the boundaries that were just described.

Malcolm Beasley followed up on Tom Arrison's comment about universities obtaining resources from foundations and other private sources. He felt it would be more appropriate to try certain experiments. It is not the intended role of any foundation to pay the continuing costs.

The Role of Publication and Communication in the Sciences

Steve Berry asked Ann Wolpert if she would accept that the role of publication still remains central to the process of communication in the sciences. Although colleagues in many different places talk more now, when it comes to the substance on which scientists base their inferences and new ideas, they have to go back to the literature. They have to go back to the publication from last month or the last decade. The scientific publication still remains the rock on which they build, and the thing that has changed most is the communication and the way scientists collaborate with each other.

Ann Wolpert agreed. However, she added that the challenge that confronts us now is the migration of the official version of the publication of record from print to electronic formats. For 15 years, people have been asking when the library will become entirely digital. The answer is, it depends, like the law. But as a practical matter, we are in a serious transition phase right now. She has no idea how long this is going to last, but we will know that we have successfully tipped into the electronic environment and learned how to archive material that is formatted digitally in reliable and sustainable ways when someone gets a Nobel Prize based entirely on an electronic publishing record. That is her standard.

Steve Berry noted that it has sometimes been said that paper is much more permanent than any current electronic form, and there is the dilemma of transferring last year's electronic archive into next year's electronic archive. However, there was a period when *Physical Review* was printed on very high acid paper; if one opens a mid-1970's *Physical Review* now, the pages crack and come apart. APS was able to transfer those fragile paper versions to electronic form and save the records in PROLA.

University Libraries and Publishing Deals

Robert Bovenschulte asked Ann Wolpert if he could conclude from her remarks that she might roughly divide all library deals into three categories: good deals, acceptable deals, and bad deals. He was curious to understand better the interplay of forces within the university when a library accepts a bad deal.

Ann Wolpert noted that there are different flavors of bad deals. Some of them have to do with the cost of the deal, and some of them have to do with the terms and conditions of use that come with the deal. A bad deal for a university is a deal that over time consistently favors the needs and interests of one group within the university over others. Part of the political risk that the scientific community runs right now, and where there is push-back in the university environment, is around the constant percentage increase in demand to support the scientific and technical literature out of a finite amount of money. At the end of the day, someone on the campus gets shortchanged as a consequence of the need to constantly feed a set of growing expectations about the payout from university library budgets in support of scientific and technical literature.

From the standpoint of the university, ultimately the groups that are shortchanged will have their time in the spotlight. Disciplines cycle through favor. Around the turn of the last century it was mechanical engineering then civil engineering and then physics; right now it happens to be the biological sciences. But as a practical matter, sometime they will cycle out too, and something else will replace them. Libraries struggle to maintain a balance on their campuses from the expenditure point of view between and among the disciplines. The potential for long-term damage is there, because one cannot buy a book that is not available anymore.

The other worry that universities have is about the terms that licenses about who can use material and under what conditions. That sometimes disadvantages parts of the campus. A library can afford to license for a limited number of users or for one subset of its community, which disadvantages those who cannot easily get to use it. Those licenses cannot be networked, or they can only be networked within a particular subset of buildings on campus. So these are the kinds of complications that libraries did not have 10 or 15 years ago, which they now confront on a regular basis.

Retrospective Information in an Electronic Environment

Gordon Neavill, of Wayne State University, noted that one of the problems in the digital environment is that the economic link between current and retrospective information is broken. In the print environment, almost all the information that libraries bought was acquired because it was current. It was then simply retained and became valuable retrospective information. In the electronic environment, we pay once for current information, then we probably have to pay all over again, at a fairly high cost, to capture the same information for very important, but low-level, retrospective uses. In the case of electronic databases, snapshots are made, but fewer people will use the retrospective information than need the current information. To some extent, these retrospective costs can be shared. He asked if electronic systems can be designed to minimize the additional cost required to retain them for retrospective purposes.

Ann Wolpert said that there are two ways to think about that. When only one format existed, it served a variety of needs: the current information dissemination, the near-term research requirement, and the long-term archiving requirement. In the digital environment, people want to use the digital materials for the sake of convenience and productivity, but there are difficulties with that, in that the electronic material does not perfectly mirror what came out in print. We heard previously that a third of the use of the Reed Elsevier titles was for nonarticle material. So there is a lot of information in print publications that is not in electronic collections of journals.

The other way to think about the electronic environment is that there are presumably costs associated with buying the retrospective collection on an annual basis. If a library had a choice of buying one year's worth at a time in terms of budget, it could do that and stop. But in the electronic environment, a library has to buy the current year plus the archive in many instances over and over again. Certainly that is the model for reference books. If a library had limited funds, it bought a scientific and technical

103

reference book one year and then bought it on an every other year or every third year basis. If this reference book is purchased electronically, the same price is paid year after year. So it affects the economics of how a library thinks about its collections, not just in terms of the current material, but in terms of how it manages the archive.

Jim O'Donnell added that the retrospective collection has never been free. It costs a lot of money to keep it dry and cool and to reshelve it. Universities spend immense amounts of money on redundant collections. Where off-site shelving is beginning to be done cooperatively among institutions, it certainly provides an alternative opportunity to think about just how much redundancy is worth paying for in terms of the use they get out of it. Universities have to be rational about how much they spend on the print archives—and it is a lot.

His comment reminded Ann Wolpert that that there are different kinds of money in universities. Capital funds are needed to build a building, which is a big one-time effort. Although the cost of the materials that are stored there over the life of the building are amortized, in fact, for the university the economics work differently. Costs in the electronic environment are annual operating costs as opposed to being able to move some of those costs off into a capital budget, and manage them differently.

CLOSING REMARKS

Ted Shortliffe

Ted Shortliffe closed the symposium by thanking the speakers and participants. One of the charges to the steering committee is to take the lessons of the symposium and try to crystallize them, and in particular to ask what are the key potential areas of study that the National Academies might focus on for additional work to help move these issues forward. There are many potential topics for study, and he welcomed advice from symposium participants on which areas the National Academies could make useful contributions in, such as areas that have not been looked at effectively by others perhaps, and where there is a need for further work. On that note, he thanked the National Academies staff for assisting with the symposium and adjourned the meeting.

Appendix A

Symposium Agenda

Symposium on Electronic Scientific, Technical, and Medical Journal Publishing and Its Implications

Main Auditorium
The National Academies
2100 C Street, NW
Washington, DC 20418

Monday, May 19

8:00	*Registration and continental breakfast*
8:30	Welcoming Remarks *Bruce Alberts, President, National Academy of Sciences*
8:45	Symposium Overview *Edward Shortliffe, professor and chair, Department of Medical Informatics deputy vice president for Information Technology Health Sciences Division, Columbia University and symposium chair*
9:00	Keynote Address *James Duderstadt, president emeritus and University Professor of Science and Engineering Millennium Project, University of Michigan*

Panel 1: Costs of Publication
Moderator: Floyd Bloom, The Scripps Research Institute

9:30	Opening Remarks by Moderator
9:35	Overview Presentation *Michael Keller, CEO, HighWire Press*

9:55	Comments by panel participants
	    Kent Anderson, publishing director, *New England Journal of Medicine*
	    Robert Bovenschulte, director, Publications Division, American Chemical Society
	    Bernard Rous, deputy director/electronic publisher, Association for Computing Machinery
	    Gordon Tibbitts, president, Blackwell Publishing USA

10:25	*Break*
10:45	Discussion of Issues
12:00	*Lunch*

105

Panel 2: Publication Business Models and Revenue
Moderator: Jeffrey MacKie-Mason, University of Michigan

1:00   Opening Remarks by Moderator

1:05   Comments by Panel Participants
    Joseph Esposito, president and CEO, SRI Consulting
    Wendy Lougee, director, University of Minnesota Library
    Brian Crawford, vice president and general manager, Life and Medical Sciences,
    John Wiley & Sons
    Patrick Brown, professor of biochemistry, Stanford University

1:55   Discussion of Issues

3:10   *Break*

Panel 3: Legal Issues in Production, Dissemination, and Use
Moderators: Ann Okerson, Yale University, and Jane Ginsburg, Columbia Law School

3:30   Copyright Basics: Ownership and Rights
*Jane Ginsburg, Morton L. Janklow Professor of Literary and Artistic Property Law,
Columbia Law School*

3:50   Economic and Non-Economic Rewards to Authors
*Michael Jensen, Jesse Isidor Straus Professor of Business Administration, Emeritus,
Harvard Business School*

4:10   Licensing
*Ann Okerson, associate university librarian for collections and technical services, Yale
University*

4:30   Discussion of Issues

5:45   *Adjourn*

6:00   *Reception, National Academy of Sciences' Great Hall*

Tuesday, May 20

8:00   *Continental Breakfast*

Panel 4: What Is Publishing in the Future?
Moderator: Daniel Atkins, University of Michigan

8:30   Opening Remarks by Moderator

8:35   Institutional Repositories
*Hal Abelson, Class of 1922 Professor of Computer Science and Engineering,
Massachusetts Institute of Technology*

8:50	Preprint Servers and Extensions to Other Fields
	Richard Luce, research library director, Los Alamos National Laboratory

9:05	Implications of Emerging Recommender and Reputation Systems
	Paul Resnick, associate professor, University of Michigan

9:20	Discussion of Issues

10:35	*Break*

Panel 5: What Constitutes a Publication in the Digital Environment?
Moderator: Clifford Lynch, Coalition for Networked Information

10:55	Opening Remarks by Moderator

11:00	Signal Transduction Knowledge Environment
	Monica Bradford, executive editor, Science

11:15	Publishing Large Data Sets in Astronomy—The Virtual Observatory
	Alex Szalay, Alumni Centennial Professor, Department of Physics and Astronomy, The Johns Hopkins University

11:30	Data Curation and Integration with the Literature
	David Lipman, director, National Institutes of Health/National Center for Biotechnology Information

11:45	Discussion of Issues

1:00	*Lunch*

Panel 6: Wrap-Up Session
Moderator: Mary Waltham, publishing consultant

1:55	Opening Remarks by Moderator

2:00	Symposium Summaries
	Malcolm Beasley, Theodore and Sydney Rosenberg Professor of Applied Physics, Stanford University
	James O'Donnell, provost, Georgetown University
	Ann Wolpert, director of libraries, Massachusetts Institute of Technology

2:30	Discussion of Issues

3:10	Closing Remarks by Symposium Chair, *Edward Shortliffe*

3:15	*Adjourn*

107

Appendix B

Biographical Information for Speakers and Steering Committee Members

Hal Abelson is professor of electrical engineering and computer science at Massachusetts Institute of Technology (MIT) and a fellow of the Institute of Electrical and Electronics Engineers (IEEE). He is winner of several teaching awards, including the IEEE's Booth Education Award, cited for his contributions to the teaching of undergraduate computer science. Professor Abelson's research at the MIT Artificial Intelligence Laboratory focuses on "amorphous computing," an effort to create programming technologies that can harness the power of the new computing substrates emerging from advances in microfabrication and molecular biology. He is also engaged in the interaction of law, policy, and technology as they relate to societal tensions sparked by the growth of the Internet, and he is active in projects at MIT and elsewhere to help bolster our intellectual commons. Professor Abelson is a founding director of the Free Software Foundation and a founding director of Creative Commons. He also serves as consultant to Hewlett-Packard Laboratories. At MIT, Professor Abelson is codirector of the MIT-Microsoft Research Alliance in educational technology and cohead of MIT's Council on Educational Technology.

Bruce Alberts, president of the National Academy of Sciences, is known for his work both in biochemistry and molecular biology, in particular for his extensive study of the protein complexes that allow chromosomes to be replicated. Dr. Alberts graduated from Harvard College and earned a doctorate from Harvard University in 1965. He joined the faculty of Princeton University in 1966 and after 10 years moved to the Department of Biochemistry and Biophysics at the University of California, San Francisco, where he became chair. He is one of the original authors of *The Molecular Biology of the Cell*, through four editions the leading advanced textbook in this important field. His most recent text, *Essential Cell Biology* (1998), is intended to present this subject matter to a wider audience. Dr. Alberts has long been committed to the improvement of science education, dedicating much of his time to educational projects such as City Science, a program that seeks to improve science teaching in San Francisco elementary schools.

Kent R. Anderson is the publishing director for the *New England Journal of Medicine*. Prior to joining the journal, he was director of medical journals at the American Academy of Pediatrics. He has been in health care publishing for more than 15 years and has worked as a writer, editor, designer, production manager, copy editor, managing editor, and publisher. He has also worked in continuing medical education, launched a half-dozen successful new titles, and contributed to numerous online publishing initiatives.

Daniel E. Atkins[*] is a professor in the School of Information and in the Department of Electrical and Computer Engineering at the University of Michigan (UM), Ann Arbor. He began his research career in the area of computer architecture and did pioneering work in high-speed computer arithmetic and parallel computer architecture. He has served as dean of the College of Engineering and more recently as the founding dean of the School of Information at UM. He is now director of the Alliance for Community Technology (ACT), an international partnership with philanthropy for research and development in the use of information and communication technology (ICT) to further the mission of educational and other nonprofit organizations. Dr. Atkins does research and teaching in the area of distributed knowledge systems. He has directed several large experimental digital library projects as well as projects to explore the application of "collaboratories" to scientific research. He has recently served as chair of the National Science Foundation Advisory Panel on Cyberinfrastructure.. The panel issued a report in February 2003 recommending a major Advanced Cyberinfrastructure Initiative intended to revolutionize science and engineering research and education. He also serves regularly on panels of the National Academies exploring issues such as scholarship in the digital age, the future of scholarly communication, and the impact of information technology on the future of higher education. He is coauthor of a recent book entitled *Higher Education in the Digital Age: Technology Issues and Strategies for American Colleges and Universities.* He serves as a consultant to industry, foundations, educational institutions, and government.

Malcolm R. Beasley is professor of applied physics in the Geballe Laboratory for Advanced Materials at

---

[*] Indicates member of the Symposium Steering Committee.

108

Stanford University. He received his B.Eng. and his PhD in physics from Cornell University. He then went to Harvard University as a research fellow and subsequently became a member of the faculty. In 1974 he joined the faculty of Stanford University where he became a full professor of applied physics in 1979. He served as the chairman of the Department of Applied Physics at Stanford from 1985-1989. In 1990 he was named the Theodore and Sydney Rosenberg Professor of Applied Physics. From 1992 to 1998 he served as director of the Center for Materials Research. And from 1998 to 2001 he served as dean of the School of Humanities and Sciences. Professor Beasley is a member of Tau Beta Pi, the IEEE, the National Academy of Sciences, and the American Academy of Arts and Sciences, and a fellow of the American Physical Society and the American Association for the Advancement of Science. He is the recipient of the Dean's Award for Superior Teaching at Stanford University. He has served as a consultant to the National Science Foundation, Department of Energy, Defense Advanced Research Projects Agency, and various industrial laboratories. He has also served on various panels of the National Research Council of the National Academies. He was an elected member of the Board of Trustees of Associated Universities, Inc., for the period 2003-2005. Professor Beasley's research interests are in materials physics with an emphasis on basic and applied superconductivity, in particular high-temperature superconducting materials and applications, and the development and application of advanced thin-film deposition techniques for complex materials.

Floyd Bloom* is chairman of the Department of Neuropharmacology at the Scripps Research Institute. He previously served as director of Behavioral Neurobiology at the Salk Institute and as chief of the Laboratory of Neuropharmacology of National Institute of Mental Health. Dr. Bloom is a member of the National Academy of Sciences and the Institute of Medicine (IOM). He has received many awards, including the Pasarow Award in neuropsychiatry and the Hermann von Helmholtz Award, as well as a number of honorary degrees from major universities. He served as editor-in-chief of *Science* magazine from 1995-2001 and currently is the president of the American Association for the Advancement of Science Board of Directors.

Robert Bovenschulte is director of the Publications Division of the American Chemical Society (ACS), which publishes journals, magazines, books, and electronic products. Prior to joining ACS in 1997, he was vice president for publishing at the Massachusetts Medical Society, owner of the *New England Journal of Medicine* and other medical publications. His career spans scholarly, professional, trade, college, and school segments of the industry. He is a frequent speaker and moderator at conferences of publishers and librarians. Mr. Bovenschulte has served as chair of the Executive Board of the International Association of Scientific, Technical, and Medical Publishers, chair of the Board of Directors of the Copyright Clearance Center, chair of the Executive Council of the Association of American Publishers' Professional and Scholarly Publishing Division, and member of the Board of Directors of the Council on Library and Information Resources.

Monica Bradford is the executive editor of the international journal *Science* (published by the American Association for the Advancement of Science). In this position she oversees the peer-review and selection of manuscripts; the copyediting and proofreading process; and the design, production, and manufacture of the print product. Over the past few years, Ms. Bradford has been heavily involved in the development of *Science* Online. In particular, she has helped create a new line of digital products referred to as online knowledge environments. *Science's* STKE and SAGE KE, the first two products in this line, are directed at research scientists. In addition, Ms. Bradford administers the APBiotech & *Science* Prize for Young Scientists in Molecular Biology. Prior to joining the staff of *Science* in 1989, Ms. Bradford worked for the Publications Division of the American Chemical Society for nine years. She holds a bachelor's degree in chemistry from St. Mary's College, Notre Dame, IN, and has done graduate work in management at the University of Maryland. Ms. Bradford was a member of the Board of Directors of the Council of Biology Editors, served as vice-chair of the Scientific Publishing Board of the American Heart Association, and is a member of the Society for Scholarly Publishing.

Patrick O. Brown is a professor of biochemistry at Stanford University School of Medicine and a Howard Hughes Medical Institute Investigator at the Stanford Medical School. He received his B.A. in chemistry, and M.D. and Ph.D. in biochemistry from the University of Chicago. He did his graduate work with Nick Cozzarelli, studying the basic molecular mechanisms of DNA topoisomerases. Following a residency in

109

pediatrics at Children's Memorial Hospital in Chicago, he began studies of the mechanism of retroviral integration as a postdoctoral fellow at the University of California, San Francisco, working with J. Michael Bishop and Harold Varmus. Dr. Brown is a member of the National Academy of Sciences. Dr. Brown's current research uses DNA microarrays and other "genomic" approaches to explore a wide range of fundamental questions in gene regulation, cell biology, physiology, development, and medicine. For the past several years he has been working to promote open, unrestricted access to scientific and scholarly publications. He is a cofounder and codirector of the Public Library of Science, a nonprofit, open-access scientific publisher.

Brian Crawford is vice president and publishing director for Global Life and Medical Sciences within the Scientific, Technical, and Medical (STM) publishing operations of John Wiley & Sons, Inc. He holds overall responsibility for coordinating the strategic development and ongoing editorial management of Wiley's international Life and Medical Sciences publishing. Experienced in subscription journal, book, and new media publishing, Dr. Crawford has spent nearly two decades within the commercial STM information industry, most recently having served in several management roles at Wiley. Before joining Wiley, he held the position of vice president and editor with Alan R. Liss, Inc. (1988-1989), a privately held book and journal publishing firm that merged to become the wholly owned Wiley-Liss, Inc. subsidiary. Mr. Crawford began his publishing career in 1985 as an acquisitions editor within the Journals Publishing Division of Academic Press, Inc. (at that time a part of Harcourt Brace Jovanovich; now a part of Reed Elsevier). Prior to entering scientific publishing, Dr. Crawford was active in both scientific research and teaching. He had the distinction of being the first biologist to be appointed as a J. Robert Oppenheimer Fellow of the Los Alamos National Laboratory, where he was a member of the scientific research staff within the Genetics Group of the Life Sciences Division from 1981-1984, and helped to launch the Department of Energy (DOE)-sponsored human genome project. He received his Ph.D. in the biochemical and biophysical sciences at the Johns Hopkins University School of Public Health, specializing in cellular and molecular aspects of cancer genetics, under sponsorship from the National Cancer Institute. He received his B.S. cum laude in chemistry from the University of Maryland at College Park. Dr. Crawford is a member of the Board of Directors of the American Medical Publishers Association (AMPA; president-elect in 2002-2003) and the Executive Council of the Professional and Scholarly Publishing Division of the Association of American Publishers (PSP/AAP; 2001-present).

James J. Duderstadt is president emeritus and University Professor of Science and Engineering at the University of Michigan. He received his B.Eng. in electrical engineering from Yale University and his Ph.D. in engineering science and physics from the California Institute of Technology. Dr. Duderstadt joined the faculty of the University of Michigan in 1968 as professor of nuclear engineering. He became dean of the College of Engineering in 1981 and provost and vice president for academic affairs in 1986. He was appointed president of the university in 1988 and served in this role until July 1996. He currently holds a university-wide faculty appointment. Dr. Duderstadt's teaching and research interests have spanned a wide range of subjects in science, mathematics, and engineering, including work in areas such as nuclear energy, lasers, computers, science policy, and higher education. During his career, Dr. Duderstadt has received numerous awards for his research, teaching, and service activities, including the National Medal of Technology for exemplary service to the nation. Dr. Duderstadt has served on and chaired numerous public and private boards, including the National Science Board and the National Academy of Engineering. He also serves as a director of Unisys and CMS Energy. He currently chairs several major national study commissions in areas including federal research policy, higher education, information technology, and nuclear energy.

Joseph J. Esposito is president and CEO of SRI Consulting, the leading publisher of syndicated research for the global chemical industry. Over the course of his career, Mr. Esposito has been associated with various publishers in all segments of the industry and was involved from an early time with new media publishing. He has served as an executive at Simon & Schuster and Random House, as president of Merriam-Webster, and CEO of *Encyclopaedia Britannica*, where he was responsible for the launch of the first Internet service of its kind. Mr. Esposito has also worked extensively in the technology industry as a consultant, with such clients as Microsoft and Hewlett-Packard, and formerly ran the Internet communications company Tribal Voice. His primary area of concentration is the development of strategy and business models for the dissemination of digital content. He has participated in numerous trade

110

shows and has written extensively in trade magazines and journals (see, for example, the essay "The Processed Book" in the March 2003 issue of *FirstMonday* at www.firstmonday.org). Mr. Esposito is currently researching new economic models for a post-copyright age.

Jane Ginsburg* is Morton L. Janklow Professor of Literary and Artistic Property Law at Columbia Law School. Ms. Ginsburg received her J.D. from Harvard Law School and her Doctor of Law from the Universite de Paris II. She served as clerk to Judge John J. Gibbons, U.S. Court of Appeals for the Third Circuit and spent three years in private practice before turning to teaching. Her principal areas of interest and expertise are in intellectual property, comparative law, private international law, and legal methods. She has published extensively on copyrights and intellectual property and serves on the editorial boards of several intellectual property journals in the United States and abroad.

Michael C. Jensen is the managing director of the Organizational Strategy Practice at the Monitor Company and Jesse Isidor Straus Professor of Business Administration emeritus of the Harvard Business School. Professor Jensen joined the faculty of the Harvard Business School in 1985. In 1999, he left Harvard to assume his current position at the Monitor Company. He was LaClare Professor of Finance and Business Administration at the William E. Simon Graduate School of Business Administration, University of Rochester, from 1984-1988, professor from 1979-1984, associate professor from 1971-1979, and assistant professor from 1967-1971. He founded the Managerial Economics Research Center at the University of Rochester in 1977 and served as its director until 1988. Professor Jensen earned his Ph.D. in economics, finance, and accounting, and his M.B.A. in finance from the University of Chicago, and an A.B. degree from Macalester College. He also has been awarded several honorary degrees and served as a Visiting Scholar at the Tuck School of Business at Dartmouth College from July 2001 to June 2002. Professor Jensen is the author of more than 50 scientific papers, in addition to numerous articles, comments, and editorials published in the popular media on a wide range of economic, finance, and business-related topics. He is author of *Foundations of Organizational Strategy* (Harvard University Press, 1998) and *Theory of the Firm: Governance, Residual Claims, and Organizational Forms* (Harvard University Press, 2000). He is editor of *The Modern Theory of Corporate Finance* (with Clifford W. Smith, Jr., McGraw-Hill, 1984) and *Studies in the Theory of Capital Markets* (Praeger Publishers, 1972). He founded the *Journal of Financial Economics*, one of the top two scientific journals in financial economics, in 1973, serving as managing editor from 1987 to 1997, when he became founding editor. In 1990, he was named "Scholar of the Year" by the Eastern Finance Association and one of the "Year's 25 Most Fascinating Business People" by *Fortune* magazine. He is the recipient of a 1989 McKinsey Award, the 1984 Joseph Coolidge Shaw S.J. Medal by Boston College, and was awarded (with William Meckling) the Graham and Dodd Plaque and first Leo Melamed Prize for outstanding scholarship by business school teachers from the University of Chicago's Graduate School of Business. Dr. Jensen has served as consultant and board member to various corporations, foundations, and governmental agencies, and has given expert testimony before congressional and state committees and state and federal courts. He is past president of the American Finance Association and the Western Economic Association International and a fellow of the American Finance Association, of the American Academy of Arts and Sciences and of the European Corporate Governance Institute.

Michael A. Keller is the Ida M. Green University Librarian at Stanford University, director of academic information resources, publisher of HighWire Press, and publisher of the Stanford University Press. Formerly he has been in library leadership positions at Cornell, Berkeley, and Yale, most actively engaged in collection development with broad exposure to the global publication and bookselling trades. In 1995, in response to scholars' requests for assistance to their scholarly societies, he established the HighWire Press as an enterprise within the Stanford University Libraries to provide online copublishing services initially to three scholarly journals. As of April 2003, HighWire Press has grown to support 361 high-impact STM journals among more than 120 major scholarly societies. It is also the site creation and host service for the revolutionary, online, third edition of the *Oxford English Dictionary*. Mr. Keller is now fostering development of additional information tools and services for the scholarly community based on the successful HighWire model, such as the LOCKSS network caching application. He serves on the boards of or as an adviser to several organizations, both for-profit and not-for-profit, including the Long Now Foundation, the Digital Library Federation, the Pacific Neighborhood Coalition, and the World Economic Forum. Mr. Keller has consulted for a variety of institutions and programs, including the City of

Ferrara in Italy, *Newsweek* magazine, Princeton, Cornell, Indiana and other universities, and several information technology companies as well as some of the numerous scholarly societies whose publishing enterprises HighWire Press supports. For more information see http://highwire.stanford.edu/~mkeller/.

David Lipman is the director of the National Center for Biotechnology Information (NCBI), which is a division of the National Library of Medicine within the National Institutes of Health. NCBI was created by Congress in 1988 to do basic research in computational biology and to develop computational tools, databases, and information systems for molecular biology. After medical training, Dr. Lipman joined the Mathematical Research Branch of the National Institute of Diabetes, Digestive, and Kidney Diseases as a research fellow. In his research on computational tools, he developed the most widely used methods for searching biological sequence databases. There are thousands of citations to Dr. Lipman's methods in papers, which have used them to discover biological functions for unknown sequences and that have thereby advanced the understanding of the molecular basis of human disease. Since 1989, Dr. Lipman has been the director of the NCBI, a leading research center in computational biology and one of the most heavily used sites in the world for the search and retrieval of biomedical information.

Wendy Pradt Lougee is university librarian and McKnight Presidential Professor at the University of Minnesota (appointed June 2002). As university librarian, she is responsible for a system of 16 libraries on the Twin Cities campus. Prior to her appointment at Minnesota, Ms. Lougee served as associate director of libraries at the University of Michigan, with responsibility for digital library development. Michigan's distinction as a premier digital library enterprise developed from a number of significant efforts launched during her tenure, including JSTOR, Making of America, the PEAK (Pricing Electronic Access to Knowledge) Project, OAI harvesting initiative, and a number of publisher collaborations. Ms. Lougee holds a B.A. in English from Lawrence University, an M.S. in library science from the University of Wisconsin, and an M.A. in psychology from the University of Minnesota.

Richard E. Luce is the research library director at Los Alamos National Laboratory (LANL). He was appointed project leader of LANL's Library Without Walls in 1994, an internationally recognized pioneering large-scale digital library. Mr. Luce holds numerous advisory and consultative positions supporting digital library development and electronic publishing. In 1999 he cofounded the Open Archives Initiative to develop interoperable standards for author self-archiving systems. Currently, he is the senior adviser to the Max Planck Society's Center for Information Management, an executive board member of the National Information Standards Organization, and a member of the University of California System-wide Library and Scholarly Information Advisory Committee. He is the course director of the International Spring School on the Digital Library and E-Publishing for Science and Technology in Geneva and a founding member of the Alliance for Innovation in Science and Technology Information. Mr. Luce received a 1996 Los Alamos Distinguished Performance Award for his contributions "introducing technological innovations supporting science and technology." The research library was corecipient of the 1999 Federal Library and Information Center of the Year award and a 1997 and 2000 Quality New Mexico "Roadrunner" recipient for organizational performance excellence based on the Malcolm Baldrige criteria.

Clifford Lynch* is the director of the Coalition for Networked Information (CNI). Prior to joining CNI in 1997, Dr. Lynch spent 18 years at the University of California Office of the President, the last 10 as director of library automation. He holds a Ph.D. in computer science from the University of California, Berkeley, and is an adjunct professor at Berkeley's School of Information Management and Systems. He is a past president of the American Society for Information Science and a fellow of the American Association for the Advancement of Science and the National Information Standards Organization. Dr. Lynch currently serves on the Internet2 Applications Council and is a member of the NRC Committee to Study Digital Archiving and the National Archives and Records Administration. He was a member of the NRC committee that published *The Digital Dilemma: Intellectual Property in the Information Infrastructure* and served on the NRC committee on Broadband Last-Mile Technology.

Jeffrey MacKie-Mason* is Arthur W. Burks Professor of Information and Computer Science and a professor of economics and public policy at the University of Michigan. He is also the founding director of the Program for Research on the Information Economy at the university and the director of doctoral studies at the School of Information at Michigan. His work is primarily in information economics, especially the Internet and advanced telecommunications technologies, and the economics of digital information

112

content. Professor MacKie-Mason received his A.B. in environmental policy from Dartmouth, his Master of Public Policy from University of Michigan, and his Ph.D. in economics from MIT.

James J. O'Donnell is professor of classics and provost at Georgetown University. He served in 2003 as president of the American Philological Association and has been elected a fellow of the Medieval Academy of America. In 1990, he cofounded *Bryn Mawr Classical Review*, the second-oldest humanities e-journal. His 1998 book, *Avatars of the Word: From Papyrus to Cyberspace* (Harvard University Press), explores the impact of technologies of writing on the shaping of culture from antiquity to the present.

Ann Okerson* is associate university librarian for collections and technical services at Yale University. She is responsible for making digital collections available to the many and varied users at Yale and has become an expert in licensing digital information for academic use and on building consortia of libraries to achieve the most effective service at the best price for academic users. Prior to joining Yale, she served as director of the Office of Scientific and Academic Publishing at the Association of Research Libraries. Ms. Okerson was named Serials Librarian of the Year in 1993 by the American Library Association (ALA) and is also the 1999 recipient of their LITA/High Tech award. She was a member of the NRC Committee on Information Technology Strategy for the Library of Congress.

Paul Resnick is an associate professor at the University of Michigan School of Information. He previously worked as a researcher at AT&T Labs and AT&T Bell Labs, and as an assistant professor at Massachusetts Institute of Technology's Sloan School of Management. He received his master's and Ph.D. degrees in electrical engineering and computer science from MIT, and a bachelor's degree in mathematics from the University of Michigan. Professor Resnick's research focuses on sociotechnical capital, productive social relations that are enabled by the ongoing use of information and communication technology. He was a pioneer in the field of recommender systems (sometimes called collaborative filtering or social filtering). Recommender systems guide people to interesting materials based on recommendations from other people. His current research focuses on reputation systems, which apply the ideas of recommender systems to evaluating people.

Bernard Rous is the deputy director of publications and electronic publishing program director for the Association of Computing Machinery. He received his undergraduate education at Brandeis University, and an M.A. at the New School for Social Research in anthropology. Mr. Rous has worked in publishing at ACM from 1980 to the present. His responsibilities have included development and management of a database publishing system for reference publications; development of early CD-ROM and hypertext products; project manager for SGML publishing production system; development and direction of electronic publishing program; drafting copyright and permissions policy for the networked environment; and establishment of Digital Library, Online Bibliographic Database, and Computing Portal with appropriate business models.

Edward H. Shortliffe* (*chair, Symposium Steering Committee*) is a professor and chair of the Department of Medical Informatics and deputy vice president for information technology for the Health Sciences Division of Columbia University. His research interests include medical informatics; issues related to integrated decision-support systems and their effective implementation; clinical medicine; and medical-informatics research and teaching. Prior to his current position, he was at Stanford University. Dr. Shortliffe provides expertise in both medicine and computer science. He received an A.B. from Harvard in applied mathematics, a Ph.D. from Stanford in medical information sciences, and an M.D. from Stanford. Dr. Shortliffe is a member of the Institute of Medicine (IOM), the IOM Council, and the NRC Committee on Science, Engineering, and Public Policy. He also served as chair of the NRC Committee on Enhancing the Internet for Biomedical Applications: Technical Requirements and Implementation Strategies.

Alexander Szalay is the Alumni Centennial Professor of Astronomy and professor of computer science at the Johns Hopkins University. He is a cosmologist, working on the statistical measures of the spatial distribution of galaxies and galaxy formation. He is the architect for the Science Archive of the Sloan Digital Sky Survey. Professor Szalay is project director of the National Science Foundation (NSF)-funded National Virtual Observatory.

Gordon Tibbitts is president of Blackwell Publishing, USA. Since 1982, he has worked to integrate

113

electronic publishing technology into traditional publishing business. Early on, he recognized the advantages of automating prepress and manufacturing work, including editorial, production, and typesetting functions, and successfully integrated the entire digital work flow at Aster Publishing Corporation in the mid-1980s. He also was responsible for automating publishing processes, creating SGML/XML/HTML-encoded content, and delivering products electronically in the early 1990s for American Health Consultants (a Thomson company). In addition to his publishing experience, Mr. Tibbitts worked for several years leading health care software development for two other Thomson information technology holdings, DKD and The Medstat Group. He developed systems sold to leading hospitals and clinics in the United States, integrating clinical and financial informatics, applying clinical best practice with performance optimization concepts drawn from evidence-based, disease-staging, case management, and JIT fields of learning. Throughout his career, Mr. Tibbitts has driven organizations toward the leading edge of technology's application in disseminating information. With Blackwell, he plays a major role in the development of all corporate technology initiatives, including Web-based content development. He has worked in executive positions for Aster Publishing, Advanstar, and The Thomson Corporation. Mr. Tibbitts received a B.S. in computer science and an M.B.A. from the University of Oregon.

Mary Waltham* is a publishing consultant. She was most recently president and publisher for *Nature* and the Nature family of journals in the United States, and formerly managing director and publisher of *The Lancet* in the United Kingdom. She founded her own consulting company four years ago. Its purpose is to help international scientific, technical, and medical publishers to confront the rapid change that the networked economy poses to their traditional business models, and to develop the new opportunities to build publications that deliver outstanding scientific and economic value. Ms. Waltham has worked at a senior level in science and medical publishing companies across a range of media, which include textbooks, magazines, newsletters, journals, and open learning materials. She served on the NRC Committee on Community Standards for Sharing Publication-Related Data and Materials.

Ann Wolpert became director of libraries for the Massachusetts Institute of Technology in January 1996. She oversees this distributed library system, which consists of five major collections, five smaller branch libraries in specialized subject areas, a fee-for-services group, and the Institute Archives. As of January 1999, her position expanded to include reporting responsibility for the MIT Press, which publishes approximately 200 new books and more than 40 journals per year in fields related to science and technology. Recently, Ms. Wolpert also assumed oversight of *Technology Review*, MIT's magazine of innovation. Ms. Wolpert's institute responsibilities include membership on the Committee on Copyright and Patents, the Council on Educational Technology, the Campus Plan Steering Committee, the Deans' Committee, and the President's Academic Council. She chairs the Management Board of the MIT Press, serves on the OpenCourseWare Interim Management Board, and is cochair of the Internal Review Committee for Financial Systems Services and Information Systems. Prior to joining MIT, Ms. Wolpert was executive director of library and information services at the Harvard Business School. Her experience also includes management of the Information Center of Arthur D. Little, Inc., where she additionally engaged in consulting assignments. More recent consulting assignments have taken her to Adelphi University in New York, to the campuses of INCAE in Costa Rica and Nicaragua, and to the Malaysia University of Science and Technology, Selangor, Malaysia. Ms. Wolpert is active in the professional library community. She currently serves on the Association of Research Libraries (ARL) Board of Directors and is a member of ARL's Scholarly Communication Committee and of its Internet2 Working Group. She also serves on the Board of Directors of the Boston Library Consortium. In addition, she is a member of the editorial boards of *Library & Information Science Research* and *The Journal of Library Administration*. A frequent speaker and writer, she has recently contributed papers on such topics as library service to remote library users, intellectual property management in the electronic environment, and the future of research libraries in the digital age. Ms. Wolpert serves on the Board of Trustees of Simmons College. In 1998 she was elected to the National Network for Women Leaders in Higher Education of the American Council on Education. She received a B.A. from Boston University and an M.L.S. from Simmons College.

114

Appendix C

Symposium Participants

Laura Abate
Himmelfarb Library, The George Washington University
leabate@gwu.edu

Hal Abelson
Massachusetts Institute of Technology
hal@mit.edu

Karen Albert
Talbot Research Library, Fox Chase Cancer Center
km_albert@fccc.edu

Bruce Alberts
National Academy of Sciences
balberts@nas.edu

Kerri Allen
Association of Research Libraries
kerri@arl.org

Mohammad Al-Ubaydli
National Center for Biotechnology Information (NCBI)
mo@mo.md

Kent Anderson
New England Journal of Medicine
kanderson@nejm.org

Carl Anderson
Drexel University Libraries
ca25@drexel.edu

Tom Arrison
The National Academies
tarrison@nas.edu

Susan Ashworth
Glasgow University
s.ashworth@lib.gla.ac.uk

Daniel Atkins
University of Michigan
atkins@umich.edu

John Aubry
American Museum of Natural History Library
jaubry@amnh.org

Midori Baer-Price
INFORMS, The Institute for Operations Research and the Management Sciences
midori.baer-price@informs.org

Marcus Banks
National Library of Medicine
BanksM@mail.nlm.nih.gov

Lori Barber
ScholarOne
lori.barber@scholarone.com

Edward Barnas
Cambridge University Press
ebarnas@cup.org

Svetla Baykoucheva
American Chemical Society (ACS)
s_baykouchev@acs.org

Malcolm Beasley
Stanford University
beasley@stanford.edu

Philippa Benson
CABS, Conservation International
p.benson@conservation.org

Sandra Berlin
American Anthropological Association
sberlin@aaanet.org

R. Steven Berry
The University of Chicago
berry@uchicago.edu

Iona Black
Yale University, Dept. of Chemistry
iona.black@yale.edu

Floyd Bloom
The Scripps Research Institute
fbloom@scripps.edu

Ronald Bluestone
Library of Congress
rblu@loc.gov

Martin Blume
The American Physical Society
blume@aps.org

115

Robert Bovenschulte
American Chemical Society Publications
rbovenschulte@acs.org

Monica Bradford
American Association for the Advancement of
Science (AAAS)
mbradfor@aaas.org

Kerry Brenner
The National Academies, BLS
kbrenner@nas.edu

Sarah Brookhart
American Psychological Society
sbrookhart@psychologicalscience.org

Patrick Brown
Stanford University
pbrown@cmgm.stanford.edu

David Bruggeman
The National Academies
dbruggem@nas.edu

Edward Campion
New England Journal of Medicine
ecampion@nejm.org

Gene Carbona
The Medical Letter
gene@medicalletter.org

Bonnie Carroll
CENDI
bcarroll@infointl.com

Mary Case
Association of Research Librarians
marycase@arl.org

Meredith Cawley
AIAA
meredithc@aiaa.org

Guy Chalk
Johns Hopkins University Center for
Communication Programs
gchalk@jhuccp.org

Leslie Chan
University of Toronto at Scarborough
chan@utsc.utoronto.ca

Nina Chang
Elsevier
n.chang@elsevier.com

Bonnie Chojnacki
University of Maryland
bc128@umail.umd.edu

Pamela Clapp Hinkle
Marine Biological Laboratory
pclapp@mbl.edu

Jeff Clark
James Madison University
clarkjc@jmu.edu

William Cohen
The Biological Bulletin, Hunter College
cohen@genectr.hunter.cuny.edu

Eileen Collins
S&T Studies
el.collins@verizon.net

Jim Comes
Univ. of Massachusetts Medical School
james.comes@umassmed.edu

Joan Comstock
Cadmus
comstockj@cadmus.com

Bridget Coughlin
The National Academies
bcoughli@nas.edu

C. Arleen Courtney
American Chemical Society
a_courtney@acs.org

Denise Covey
Carnegie Mellon, Hunt Library
troll@andrew.cmu.edu

Nicholas Cozzarelli
University of California, Berkeley
ncozzare@socrates.Berkeley.edu

Brian Crawford
John Wiley & Sons, Inc.
bcrawfor@wiley.com

Vicki Croft
Washington State University
croft@wsu.edu

116

Raym Crow
Chain Bridge Group
crow@chainbridgegroup.com

Bruce Dancik
NRC Research Press c/o Renewable Resources
bruce.dancik@ualberta.ca

Lloyd Davidson
Northwestern University, Mudd Library for
Science & Engineering
Ldavids@northwestern.edu

Adrienne Davis
The National Academies
adavis@nas.edu

Bart De Castro
Cambridge Scientific Abstracts
bdecastro@csa.com

Stephanie Dean
American Society for Cell Biology (ASCB)
sdean@ascb.org

Laura Dean
National Center for Biotechnology Information
laura@idiopathic.com

Masako Dickinson
American Chemical Society
med96@acs.org

Heather Dittbrenner
Independent Consultant
hdittbrenner@comcast.net

Lisa Dittrich
Association of American Medical Colleges
lrdittrich@aamc.org

Richard Dodenhoff
American Society for Pharmacology and
Experimental Therapeutics
rdodenhoff@aspet.org

Daniel Dollar
Yale University, Cushing/Whitney Medical
Library
daniel.dollar@yale.edu

Mark Doyle
The American Physical Society
doyle@aps.org

Guy Dresser
Allen Press, Inc.
gdresser@allenpress.com

Michael Droettboom
Johns Hopkins University
mdboom@jhu.edu

James Duderstadt
University of Michigan
jjd@umich.edu

Carol Edwards
American Statistical Association
caroled@amstat.org

Anita Eisenstadt
National Science Foundation
aeisenst@nsf.gov

Julie Esanu
The National Academies, BISO
jesanu@nas.edu

Joseph Esposito
SRI Consulting
jesposito@sric.sri.com

Walter Finch
National Technical Information Service
wfinch@ntis.gov

Henry Flores
Copyright Clearance Center, Inc.
hflores@copyright.com

Martin Frank
American Physiological Society
mfrank@the-aps.org

Mark Frankel
AAAS
mfrankel@aaas.org

Tracie Frederick
Georgetown University Medical Center
tef7@georgetown.edu

Theodore Freeman
Allen Press, Inc.
tfreeman@allenpress.com

Amy Friedlander
Council on Library and Information Resources
amfr@loc.gov

117

Fred Friend
Joint Information Systems Committee
f.friend@ucl.ac.uk

Sherrilynne Fuller
University of Washington
sfuller@u.washington.edu

Ken Fulton
The National Academies
kfulton@nas.edu

John Gardenier
Independent Consultant
drgarden@erols.com

Lorrin Garson
American Chemical Society
lgarson@acs.org

Julia Gelfand
University of California, Irvine Libraries
jgelfand@uci.edu

Jane Ginsburg
Columbia University School of Law
ginsburg@law.columbia.edu

Carter Glass
American Geophysical Union
cglass@agu.org

Erica Goldstein
American Society for Nutritional Sciences
goldsteine@asns.org

Barbara Gordon
American Society for Biochemistry & Molecular
Biology
bgordon@asbmb.faseb.org

Albert Greco
Fordham University
angreco@aol.com

Michael Greenberg
The Whitney Laboratory
mjgberg@aug.com

Suzanne Grefsheim
NIH Libarary
sg8d@nih.gov

Jane Griffith
National Library of Medicine
jbgriffith@nlm.nih.gov

Anne Gwynn
Library of Congress, Congressional Research
Services
agwynn@crs.loc.gov

Melissa Hagemann
Open Society Institute
mhagemann@sorosny.org

Joel Hammond
BIOSIS
jkhammond@biosis.org

Charles Hancock
American Society for Biochemistry and
Molecular Biology (ASBMB)
chancock@asbmb.faseb.org

Holly Harden
Welch Medical Library, Johns Hopkins Medical
Institutions
hharden@mail.jhmi.edu

Elizabeth Have
BIOSIS
btenhave@biosis.org

Michael Held
The Rockefeller University Press
held@rockefeller.edu

Stephen Heller
National Institute of Standards and Technology
(NIST)
srheller@nist.gov

Robert Hershey
Medical Doctor
hershey@cpcug.org

Derek Hill
National Science Foundation
dhill@nsf.gov

Carol Hunter
University of Virginia
chunter@virginia.edu

Alvin Hutchinson
Smithsonian Institution Libraries, National
Zoological Park Library
hutchinsona@si.edu

John Inglis
Cold Spring Harbor Laboratory Press
inglis@cshl.org

118

Demarie Jackson
American Psychological Association
djackson@apa.org

Beth Jacoby
University of Maryland, Baltimore, Health
Science Library
bjacoby@hshsl.umaryland.edu

Brent Jacocks
American Speech-Language Hearing
Association (ASHA)
journals@asha.org

Michael Jensen
National Academies Press
mjensen@nas.edu

Michael Jensen
Harvard University
michael_jensen@ssrn.com

Nels Jensen
Blackwell Publishing
Nels.jensen@usa.blackwellpublishing.com

Rick Johnson
SPARC
rick@arl.org

Heather Joseph
BioOne
heather@arl.org

Lisa Junker
American Industrial Hygiene Association
ljunker@aiha.org

Brian Kahin
University of Michigan, Ford School of Public
Policy
kahin@umich.edu

Neal Kaske
University of Maryland
nkaske@umd.edu

Cara Kaufman
Kaufman-Wills Group, LLC
ckaufman@bellatlantic.net

Irena Kavalek
U.S. Geological Survey Library
ikavalek@usgs.gov

Michael Keller
HighWire Press
michael.keller@stanford.edu

Maureen Kelly
Consultant
mckelly@ix.netcom.com

Donald King
University of Pittsburgh
dwking@pitt.edu

Gary Kittredge
Capital City Press, Inc.
gkittred@capcitypress.com

Barbara Koehler
Welch Library, Johns Hopkins
bmk@jhmi.edu

Michael Koenig
Long Island University
michael.koenig@liu.edu

Sheldon Kotzin
National Library of Medicine/NIH
ann_bornstein@nlm.nih.gov

Alan Kraut
American Psychological Society
akraut@psychologicalscience.org

Marc Krellenstein
Reed Elsevier
m.krellenstein@elsevier.com

Thomas Kuhn
American College of Physicians
tkuhn@acponline.org

Catherine Langrehr
Association of American Universities
catherine_langrehr@aau.edu

Judith LaVoie
VA Rehabilitation Research & Development
judith@vard.org

Ann Link
American Association of Immunologists - The
Journal of Immunology
alink@ji.faseb.org

Anne Linton
George Washington University
mlbaml@gwumc.edu

119

David Lipman
National Institutes of Health/National Center for
Biotechnology Information
dl2a@nih.gov

Wendy Lougee
University of Minnesota Library
wlougee@umn.edu

Michael Luby
Columbia University
ml1047@columbia.edu

Richard Luce
Los Alamos National Laboratory
rick.luce@lanl.gov

Richard Lucier
Dartmouth College
richard.e.lucier@dartmouth.edu

Marilyn Lux
American College of Surgeons
mlux@facs.org

Clifford Lynch
Coalition for Networked Information
clifford@cni.org

James MacDonald
The American Phytopathological Society
jdmacdonald@ucdavis.edu

Jeffrey MacKie-Mason
University of Michigan
jmm@umich.edu

Amanda Maguire
American Institute of Aeronautics and
Astronautics
amandam@aiaa.org

Gene Major
National Aeronautics and Space Administration
(NASA)
major@gcmd.nasa.gov

Constance Malpas
The New York Academy of Medicine
cmalpas@nyam.org

David Martinsen
American Chemical Society
d_martinsen@acs.org

Jan Massey
The American Association of Immunologists
jmassey@aai.faseb.org

Joseph Mazurkiewicz
Journal of Histochemistry and Cytochemistry
mazurkj@mail.amc.edu

Luke McCabe
American Institute of Aeronautics and
Astronautics
lukem@aiaa.org

Katherine McCarter
Ecological Society of America
ksm@esa.org

Johanna McEntyre
NCBI
mcentyre@ncbi.nlm.nih.gov

Bruce McHenry
Discussion Systems
bruce@discussionsystems.com

Robert McKinney
Cadmus Communications
mckinneyr@cadmus.com

Peggy Merryman
US Geological Survey Library
mmerryma@usgs.gov

Henry Metzger
The National Academies, BISO
metzgerh@exchange.nih.gov

Cynthia Middleton
U.S. Government Printing Office
cmiddleton@gpo.gov

Linda Miller
Library of Congress
lmil@loc.gov

Kenneth Miller
American Association of Colleges of Pharmacy
kmiller@aacp.org

Lenne Miller
The Endocrine Society
lmiller@endo-society.org

Kurt Molholm
Defense Technical Information Center
kmolholm@dtic.mil

Pat Molholt
Columbia University Health Sciences Division
molholt@columbia.edu

Robert Molyneux
U.S. National Commission on Libraries and
Information Science
bmolyneux@nclis.gov

Marguerite Montes
Journal of Rehabilitation Research &
Development
marx@vard.org

John Muenning
University of Chicago Press
jmuenning@press.uchicago.edu

Jane Murray
University of Maryland Baltimore, Health
Sciences & Human Services Library
jmurray@hshsl.umaryland.edu

Gordon Neavill
Wayne State University
aa3401@wayne.edu

Michael Neff
American Society for Horticultural Science
mwneff@ashs.org

Kathe Obrig
Himmelfarb Library, The George Washington
University
mlbkso@gwumc.edu

Jack Ochs
American Chemical Society
j_ochs@acs.org

James O'Donnell
Georgetown University
jod@georgetown.edu

Ann Okerson
Yale University
ann.okerson@yale.edu

Jill O'Neill
National Federation of Abstracting and
Information Services (NFAIS)
jilloneill@nfais.org

Shigeharu Ono
Kinokuniya Publications Service
ono@kinokuniya.com

David Osterbur
Harvard University
dosterbu@mcb.harvard.edu

Doris Peter
The Medical Letter, Inc.
dpeter@medicalletter.org

Walter Peter
CADMUS
peterw@cadmus.com

Stephen Phelps
Capital City Press
sphelps@capcitypress.com

Theresa Pickel
Alliance Communications Group, Allen
Press,Inc.
tpickel@allenpress.com

Kevin Pirkey
Dartmouth Journal Services
kpirkey@dartmouthjournals.com

Barbara Pope
National Academy Press
bkline@nas.edu

Heather Price
Journal of Bone & Mineral Research
heather@jbmr.org

Roberta Rand
University of Miami - Rosenthal School of
Marine & Atmospheric Science
rrand@rsmas.miami.edu

Alan Rapoport
National Science Foundation
arapopor@nsf.gov

Howard Ratner
Nature Publishing Group
h.ratner@natureny.com

Tovah Reis
Brown University
tovah_reis@brown.edu

Paul Resnick
University of Michigan
presnick@umich.edu

121

Leonard Rhine
University of Florida Health Science Center
Libraries
lenny@library.health.ufl.edu

Susan Riedel
Library of Congress
wrie@loc.gov

Nancy Rodnan
FASEB
nrodnan@faseb.org

Linda Rosenstein
University of Pennsylvania Biomedical Library
rosenstl@mail.med.upenn.edu

Beth Rosner
AAAS
Rosner@aaas.org

Bernard Rous
Association for Computing Machinery
rous@hq.acm.org

Kevin Rowan
The National Academies, COSEPUP
krowan@nas.edu

Lucia Ruggiero
World Health Organization
ruglucia@paho.org

John Rumble
NIST
john.rumble@nist.gov

John Sack
Stanford University
sack@stanford.edu

Jennifer Samuels
AIAA
jens@aiaa.org

Agnes Schonbrunn
University of Texas - Houston
agnes.schonbrunn@uth.tmc.edu

Roger Schonfeld
The Andrew W. Mellon Foundation
rcs@mellon.org

Edward Shortliffe
Columbia University
shortliffe@dmi.columbia.edu

Elliot Siegel
National Library of Medicine
siegel@nlm.nih.gov

Pamela Sieving
National Institutes of Health Library
ps256e@nih.gov

William Silberg
Medscape
bsilberg@webmd.net

Natalie Simone-Fountaine
Analytical Sciences, Inc.
nfountaine@asciences.com

Robert Simoni
Department of Biological Sciences
rsimoni@asbmb.faseb.org

Susan Skomal
American Anthropological Association
sskomal@aaanet.org

Eric Slater
American Chemical Society
e_slater@acs.org

F. Hill Slowinski
Worthington International
hslowinski@worthingtoninternational.com

MacKenzie Smith
MIT Libraries
kenzie@mit.edu

Kent Smith
National Library of Medicine
kent_smith@nlm.nih.gov

Michael Smolens
3BillionBooks
michael@3billionbooks.com

John Sopka
National Technical Information Services
wfinch@ntis.gov

Beth Staehle
American Society of Plant Biologists
beths@aspb.org

Robert Stanley
University of Alabama, Birmingham
rstanley@uabmc.edu

122

Diane Sullenberger
The National Academies
sullenb@nas.edu

Alex Szalay
Johns Hopkins University
szalay@jhu.edu

Herbert Tabor
American Society for Biochemistry & Molecular
Biology
htabor@asbmb.faseb.org

Marian Taliaferro
Association of American Medical Colleges
mtaliaferro@aamc.org

Patricia Thibodeau
Duke University Medical Center
thibo001@mc.duke.edu

Gordon Tibbitts
Blackwell Publishing
GTibbitts@bos.blackwellpublishing.com

Jeanne Slater Trimble
American Institute of Aeronautics & Astronautics
jeannet@aiaa.org

Ian Tuttle
Georgetown University
tuttlei@georgetown.edu

Paul Uhlir
The National Academies, BISO
puhlir@nas.edu

Victor Van Beuren
American Association of Pharmaceutical
Scientists
vanbeurenv@aaps.org

Pamela Van Hine
American College of Obstetricians and
Gynecologists
pvanhine@acog.org

John Vaughn
Association of American Universities
john_vaughn@aau.edu

Jack Verna
Analytical Sciences, Inc.
jverna@asciences.com

Philip Wallas
EBSCO Information Services
pwallas@ebsco.com

Mary Waltham
Independent Consultant
mary@marywaltham.com

Walt Warnick
U.S. Department of Energy
walter.warnick@science.doe.gov

Donald Waters
The Andrew W. Mellon Foundation
pam@mellon.org

Judy Weislogel
Elsevier
j.weislogel@elsevier.com

Nanette Welton
University of Washington
nwelton@u.washington.edu

Louise Wides
Wides Consulting, LLC
widesconsult@attglobal.net

Sophie Wilkinson
American Chemical Society
s_wilkinson@acs.org

Alma Wills
Kaufman-Wills Group, LLC
almawills12@comcast.net

Bonita Wilson
Corporation for National Research Initiatives
(CNRI)
bwilson@cnri.reston.va.us

Nancy Winchester
American Society of Plant Biologists
nancyw@aspb.org

Terry Wittig
George Mason University
twittig@gmu.edu

Ann Wolpert
Massachusetts Institute of Technology
awolpert@mit.edu

Melissa Yorks
National Library of Medicine
yorks@nih.gov

123

Nevenka Zdravkovska
Georgetown University
nevenkaz@georgetown.edu

Laura Zimmerman
Fry Communications, Inc
lzimmerm@frycomm.com

Eric Zimmerman
The Research Authority of Bar-Ilan University
zimmee@mail.biu.ac.il