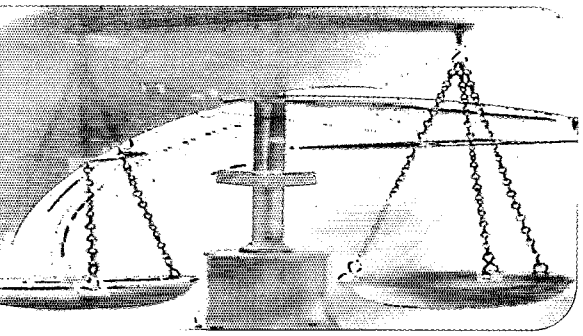


by Carol Ebbinghouse
Law Librarian
California Second District Court of Appeal
Los Angeles, CA

Open Access: The Battle for Universal, Free Knowledge

It's a war out there, skirmishes everywhere. Even within the forces for open and free access, alliances have begun to blur. Some players fight on two fronts. Some publishers are enabling authors to self-archive[1] and contribute to institutional repositories with pre- and/or post-publication, full-text versions of their articles. Search engines have entered the fray, bringing their technological resources to make open access content useful and responsive. A whole new vocabulary is emerging, along with color-coding applied to publishers and their policies. Even as I make these keystrokes, a debate rages about the use of gold, bright green, pale green, blue, yellow, and white to denote different levels of open access!

Stereotyped roles have changed. No longer does the situation involve only "black-hat" greedy journal or database publishers vs. "white hat" librarians with budgets that can't keep up with last year's subscription costs — much less this year's increases. Many publishers are joining authors in permitting open access through self-archiving in institutional repositories, primarily educational but sometimes commercial, opening up the content of these archives to the Web. Everyone has a valid point to communicate. And the battlefield extends across the world — from the U.S. to England[2] and beyond.[3]

The battle sites include negotiating tables where authors contract to sell their work to publishers and where librarians or their consortia[4] license information from publishers or database aggregators. Battle geography can also encompass end users scanning the terms and conditions of a commercial database. Ultimately, the players in the open access war include all of us, all consumers of knowledge, whether researchers working in a university library or science lab and possibly accessing "proprietary information" through a library contract; attorneys using individual Lexis or Westlaw contracts; businesspeople with corporate contracts; or cancer researchers tapping grant-funded content in research databases from around the world.

First we must learn the terminology of the conflict. In the "Definitions of Terms and Acronyms" sidebar on page 12, you will find a number of definitions, deciphered acronyms, and links to sites that are useful in tracking terminology. Old words such as "archive" take on new and very specific meanings that have nothing to do with books, parchment, and/or artifacts in secure buildings.

The open access literature is wide and varied. There are even competing sets of myths on Web sites for the different camps! On one side, we have the Association of American Publishers (AAP)[5] "setting the record straight about Academic Journal Publishing" and decrying as mythical arguments from such sources as the Public Library of Science, Congressional Representative Martin O. Sabo (D-MN) with his "Public Access to Science[6] Act," and Dr. Michael B. Eisen, Ph.D., with his Salon.com article[7] entitled "The Free Research Movement."

On the other side, "Open Access ***now***, *Campaigning for Freedom of Research Information*" [http://www.biomedcentral.com/openaccess/inquiry/myths/?myth=all] counters with "(Mis)Leading Open Access Myths." BMC finds its myths in the mouths of such luminaries as Crispin Davis, CEO of Reed Elsevier; John Jarvis, managing director of Wiley Europe; Richard Charkin, CEO of Macmillan; and Bob Campbell, president of Blackwell Publishing, in testimony[8] and submissions to the House of Commons throughout early 2004.

The Digitization Ground Action

Digitization is hot these days.[9] Individual publishers and societies[10] such as the American Physiological Society, the Endocrine Society, Elsevier Science Direct, etc., are joining data aggregators such as ProQuest's historical newspaper [www.proquest.com] and other retrospective series, Gale's primary[11] material, EBSCO [www.ebsco.com/home], traditional search services, HighWire

Press (which claims to host the largest repository of free, full-text, peer-reviewed articles in the world, with more than 750,000 articles freely available online), SPARC from ARL and its major research library members, and the Law Library Microform Consortium's (LLMC) Digital project [www.llmc.com/links.htm] to digitize its 50 million page images for a total of 100 million digitized pages by 2013. Even the search services are getting into the action with the new, much trumpeted Google project to digitize the entire book collections of five major research libraries[12], the Google/CrossRef Search[13] Pilot program, and Yahoo! Search's access arrangements with OAIster.[14]

How do these groups expect to recover costs, much less make a profit? Many publishers and developers use grants to fund[15] demonstration projects, while not-for-profit, membership-supported organizations (e.g., LLMC and the societies) have their own repositories. Others — such as the nonprofit Public Library of Science [www.plos.org/faq.html] and the for-profit BioMed Central [www.biomedcentral.com/info/authors/apcfaq] — charge authors.

The Open Access War Zone

Individual authors are developing their own self-archives, under the urging of such open access zealots as Stevan Harnad of the American Scientist Forum. The same organizations that pay authors their salaries often operate institutional archive repositories. Some of those repositories are run by academic libraries. Some may have their own revenue routes. For example, Michael Keller, University Librarian at Stanford University, also manages Stanford University Press and HighWire Press, which will begin to host all of the Oxford University Press[16] journals this year. Some of the institutions are university distance learning programs that apply material outside the physical confines of universities. Some universities[17] want wide dissemination of their faculty intellectual property, including peer-reviewed articles, at sites such as the eScholarhip site at California Digital Library or DSpace at MIT [www.dspace.org], while others draw users to their own repositories

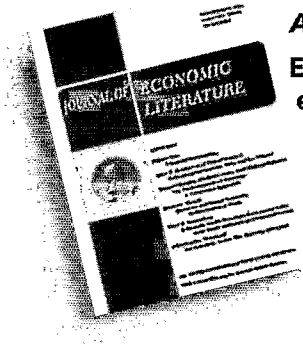
Upgrade to an Online Subscription to the *American Economic Review*, *Journal of Economic Literature*, and *Journal of Economic Perspectives*

Electronic access to the AEA journals for libraries, institutions, and firms!

Site license access includes reference links (updated daily) to full-text articles through CrossRef and Ingenta, links to full-text articles citing AEA articles, table of contents alerts, and searchable full-text. Site licenses include all years from 2002 forward so long as a subscription is active!

Contact the AEA at **www.aeaweb.org/subscribe.html** or your subscription agent to add an electronic site license to the AEA journals today!

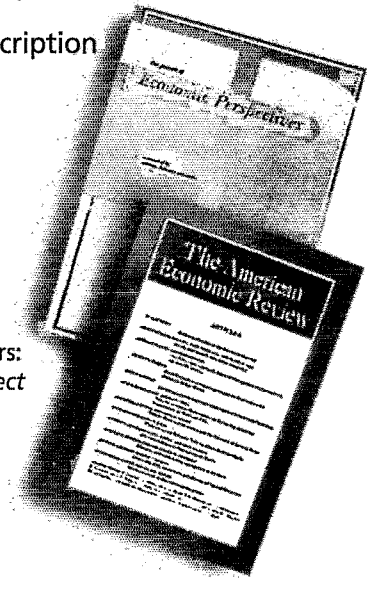
Site License price as of 1/1/2005 (includes all 3 AEA online journals).....$270.00



Ask about...

EconLit — The foremost bibliographic database in economics is linked to full-text now more than ever!

Full-text linking options are available from all EconLit vendors:
CSA - Dialog - Ebsco - OCLC - OVID/SilverPlatter - ScienceDirect



American Economic Association
2403 Sidney Street #260, Pittsburgh, PA 15203-2167, USA
info@econlit.org

using the DSpace software available on an open source[18] basis.

Search Engine Nuclear Proliferation

What does Google intend to do with all that new IPO money? Buy content, of course! The massive digitization program it has undertaken with five major research libraries could leave it with millions and millions of full-text electronic books, in and out of copyright, in and out of print. The effort might cost a half-billion dollars and take 10 years to complete, but even if it did, the possibilities for "monetizing" the content could justify the expense. In any case, users should rejoice at the prospect of all that recorded knowledge heading online to a service that likes to give information away.

Yahoo! Search has been touring the country gathering content permissions, e.g., access to OAIster, Project Gutenberg, et al. With Yahoo! and soon Microsoft beating the bushes for more content, the "Search Wars" promise to create a whole new information world.[19]

Will these search giants want to go to a revenue-seeking model, e.g., enterprise charging, pay-per view, or subscriptions? While cost and cost-recovery models might apply to some content, open access resources ought to remain free. However, easy, global, one-search-covers-all retrieval of these digitized, full-text, open access goodies could be worth its weight in gold. In other words, these giants might not "sell" the public domain or open access content, but they could try selling the easy access of a familiar search protocol.

Would they charge users? That's not their style. Sure, we'll likely see ads for sunscreen and Deep Woods Off! on the monitor with the river raft chapters of *Huckleberry Finn*, but if the book is still free to read, how many will care? Libraries working on curtailed budgets may have to partner with these cash-rich companies truly committed to open access and free content.

On the other hand, libraries are still in the game. Look at SearchLight [http://searchlight.cdlib.org/cgi-bin/searchlight] from the California Digital Library — a one-stop search portal for information. "With SearchLight, you can search many of the databases and other resources available to CDL users — all at the same time. It can find books, journal and encyclopedia articles, and quality Internet sites. SearchLight will run the search and bring back the results. With SearchLight, you don't have to worry about finding the best databases first — it does it for you." And the "results are organized into Categories — books, journal indexes, electronic journals, electronic texts and documents, reference resources, Web directories." If you are on a University of California network, you get access to all databases to which the campuses subscribe! And it is free! What a deal! SearchLight accesses more commercial information than either Google or Yahoo! And the "quality Internet sites" on the lists don't run the risk of placement influenced by payola for priority on the search results screen.

SEARCHER
The Magazine for Database Professionals

Vice President, Content	Richard Kaser
Editor-in-Chief	Barbara Quint
Manager, Editorial Services	Lauree Padgett
Copy Chief	Deborah Poulson
Junior Copy Editor	Cassie Johnson
Vice President, Graphics and Production	M. Heide Dengler
Digital Art/Graphic Design/Illustration	Jacqueline Walter Crawford Amanda Beland
Production	Bonnie Weisbrod
Group Publisher	Bob Fernekees
Advertising Director *(U.S. and International)*	Michael V. Zarrello 609-654-6266 x132 mzarrello@infotoday.com
Vice President, Marketing & Business Development	Thomas H. Hogan, Jr.
Circulation Manager	Janeen Welsh Phone: 609-654-6266 Fax: 609-654-4309
President & CEO, Information Today, Inc.	Thomas H. Hogan
Chairman of the Board	Roger Bilboul
Vice President, Administration	John Yersak
Vice President, Information Technology	Bill Spence

Information Today, Inc.
Web site: www.infotoday.com
Copyright © 2005 by Information Today, Inc. All rights reserved.

Searcher: The Magazine for Database Professionals is published 10 times per year (January, February, March, April, May, June, July/August, September, October, November/December) by Information Today, Inc., 143 Old Marlton Pike, Medford, NJ 08055. Periodicals Postage Rates paid at Medford, NJ, and additional mailing offices. Postmaster: Send address changes to Searcher The Magazine for Database Professionals, 143 Old Marlton Pike, Medford, NJ 08055. Registered in U.S. Patent and Trademark Office (ISSN 1070-4795). Subscriptions from the U.S.: $86.95; Canada and Mexico: $111; outside North America: $118 in U.S. funds. All editorial materials should be sent to Barbara Quint, Editor, 932 11th Street, Suite 9, Santa Monica, CA 90403. *Searcher* magazine is carried full text by the following online database vendors: Dialog, Datastar, Factiva, Gale Group, LexisNexis, Westlaw, EBSCO*host*, OCLC, ProQuest, and WilsonWeb. A selection of key articles and columns also appear full-text on Information Today's Web site [http://www.infotoday.com/searcher]. Authorization to photocopy items herein is granted by Information Today, Inc., provided that the base fee of $3.50 per copy, plus 50¢ per page is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

Ground Zero — Permissions

This is where the blood may hit the wall. Ask Legg Mason about the $20 million jury verdict[20] against it for making multiple copies and "posting every issue of the *Reports* on Legg-Mason's firm-wide intranet"[21] from its single subscription to *Lowry's Reports*. Ouch.

If the publisher or scholarly society that published the article does not pay or contract for the author's copyright, what does it have? Just the old familiar first North American serial rights? Can a publisher provide open access to an author's materials unilaterally because it wants to go open access? According to the Open Access\Overview[www.earlham.edu/~peters/fos/overview.htm] by Peter Suber:

> The legal basis of OA is either the consent of the copyright holder or the public domain, usually the former. Because OA uses copyright-holder consent, or the expiration of copyright, it does not require the abolition, reform, or infringement of copyright. One easy, effective, and increasingly common way for copyright holders to manifest their consent to OA is to use one of the Creative Commons licenses. They could also compose their own permission statements and attach them to their works. When copyright holders consent to OA, what are they consenting to? Usually they consent in advance to the unrestricted reading, downloading, copying, sharing, storing, printing, searching, linking, and crawling of the full text of the work. Most authors choose to retain the right to block the distribution of mangled or misattributed copies. Some choose to block commercial re-use of the work. Essentially, these conditions block plagiarism, misrepresentation, and sometimes commercial re-use, and authorize all the uses required by legitimate scholarship, including those required by the technologies that facilitate online scholarly research. For works not in the public domain, OA always requires the copyright-holder's consent.

If authors still own the copyright in their material, could they demand the return of their articles for posting on their own self-archiving Web sites? If the publisher does not compensate them for revenue from online access, can the author demand the money "*or else*"? Can authors who fail to retain copyright, as well as those who do, give their articles to open archives?[22] What if the author depends upon income from their articles for a living? For those who never contracted-away their copyright interests, could publishers sell access to their collections without compensating these authors? Could publishers give it away? If the authors or their heirs can't be found to give permission, what then? Isn't this a *Tasini*[23] issue? ("No, not him again!?!" the industry cries.) Even the concept of a lawsuit can strike fear into the hearts of corporations.

Here's the basic problem: "How can publishers or database aggregators sell what they do not own?" The Endocrine Society[24] [http://edrv.endojournals.org/] and Association for Computing Machinery[25] (ACM) [http://www.acm.org] are digitizing entire backruns of their publications, some extending back almost a century. These two appear to be consulting with their authors in their permissions and copyright policies.

One way to get around the permission issue is to limit the digital collection to content already in the public domain. Another alternative is to merely link or "index" the full texts and permit end users to "search" various repositories through the Google interface, much like the Crossref.org[26] pilot program. Yahoo! has joined OAIster to provide links[27] to the wealth of information in ArXive and other resources. For the Law Library Microform Consortium, most all of the old and current primary law (cases, codes, regulations, etc., both state and federal) is also in the public domain. The LLMC collection includes American and international resources as page images from its microfiche collections in PDF files, while other law repositories use links to "originals" in digital formats, such as Cornell's Legal Information Institute (LII) [http://www.law.cornell.edu], the American Society of International Law (ASIL) [http://www.asil.org], or the Electronic Information System for International Law (EISIL) [http://www.eisil.org].

Some journals indicate who owns the copyright in online articles. For instance, Law Library Journal [http://www.aallnet.org/products/pub_journal.asp] provides in its *Law Library Journal* Copyright Policy: "All articles copyright [year] by the American Association of Law Libraries, except where otherwise expressly indicated. Except as otherwise expressly provided, the author of each article in this issue has granted permission for copies of that article to be made for classroom use or for any other educational purpose provided that (1) copies are distributed at or below cost, (2) author and journal are identified, and (3) proper notice of copyright is affixed to each copy."[28]

Revenge of the Librarians

What if the libraries began to compete by developing their own databases? Well, they already are! Washburn University School of Law has a library Web site [http://www.washlaw.edu/] of online legal information, as does Cornell's Legal Information Institute [http://www.law.cornell.edu/]. Competing with

Definitions of Terms and Acronyms

Budapest Open Access Initiative (BOAI)
http://www.soros.org/openaccess

The Budapest Open Access Initiative arose from a small but lively meeting convened in Budapest by the Open Society Institute (OSI) on December 1-2, 2001. The purpose of the meeting was to accelerate progress in the international effort to make research articles in all academic fields freely available on the Internet.

BOAI only seeks open access for the scientific and scholarly research texts that authors give to publishers and readers without asking for any kind of royalty or payment. As the BOAI public statement puts it, "[P]rimarily, this category encompasses ... peer-reviewed journal articles, but it also includes any unreviewed preprints that [scholars] might wish to put online for comment or to alert colleagues to important research findings." It does not include books from which their authors would prefer to generate revenue. It does not include any nonscholarly writings, such as novels or news. While the BOAI does not specifically cover donated scholarship other than peer-reviewed journal articles and preprints, it could extend quite naturally to all the writings for which authors do not expect payment. These include scholarly monographs on specialized topics, conference proceedings, theses and dissertations, government reports, and statutes and judicial opinions. [From http://www.earlham.edu/~peters/fos/boaifaq.htm#literature.]

Common Content
http://commoncontent.org/

A catalog of works licensed in the Creative Commons, available to anyone for copying or creative re-use. The catalog includes 2,844 records, with many collections containing hundreds or thousands of other works.

Copyleft
http://www.gnu.org/copyleft/gpl.html

The Copyleft movement provides licenses for the distribution of software under a General Public License (GNU-GPL) that allows users the opportunity to reproduce and redistribute software programs so long as the author does not make any restrictions on the distribution of the software to later users. It is opposed to copyright, which protects the original author/creator against unauthorized copying or redistribution of the software. The Electronic Frontier Foundation [http://www.eff.org] has a collection of copylefts.

Counter Copyright

An alternative to the exclusivity of copyright, counter-copyright invites others to use and build upon a creative work. By encouraging widespread dissemination, the counter-copyright campaign fosters a rich public domain. If you place the [cc] icon at the end of your work, you signal others that you allow them to use, modify, edit, adapt, and redistribute the work you have created. The counter-copyright is not a replacement for an actual copyright, rather, it signals that the creator is willing to share their work.

Free

According to BOAI, "'Free' is ambiguous. We mean free for readers, not free for producers. We know that open-access literature is not free (without cost) to produce. But that does not foreclose the possibility of making it free of charge (without price) for readers and users. The costs of producing open-access literature are much lower than the costs of producing print literature or toll-access online literature. These low costs can be borne by any of a wide variety of potential funders..." [http://www.earlham.edu/~peters/fos/boaifaq.htm#journals]. For

commercial information providers is not a new concept! Today, the California Digital Library [http://www.cdlib.org] has several efforts that link to its own and other sites and enable searching of publicly available databases and other Internet resources. For more information on this collective effort, see the sidebar "California Digital Library Services" on page 14.

Google's library digitization project is not the only one. Another truly massive, international partnership is the Million Book Project, an international collaboration "to digitize 1 million books and offer them free-to-read on the surface Web by 2007. Led by computer scientists and librarians at Carnegie Mellon, Million Book Project partners include universities and research institutes; for-profit and not-for-profit organizations; governments and government agencies; librarians and archivists; software developers; and commercial publishers, university presses, and scholarly associations. Areas of cooperation include collection development, copyright permission, digital registry, book acquisition and shipping, scanning, quality control, sustainability, and added-value services."[29] Approximately 100,000 books should be available by 2005. The books will be replicated on servers around the world, indexed by popular search engines, and freely available on the surface Web — accessible anywhere, any time, to anyone with an Internet connection. Any school, public, or academic library will be able to link its library catalog records to the books in the Million Book Collection. The Collection will support education, research, and lifelong learning worldwide. [For more information, click to http://www.library.cmu.edu/Libraries/MBP_FAQ.html.] The project operates off the Internet Archive.

"free" business models, BOAI provides three publications: *Guide to Business Planning for Launching a New Open Access Journal* (2nd Edition); *Guide to Business Planning for Converting a Subscription-Based Journal to Open Access* (2nd Edition); and *Model Business Plan: A Supplemental Guide for Open Access Journal Developers & Publishers* (1st Edition).

Open Access (OA)

According to Stevan Harnad, "An article is OA if its online full-text can be immediately and permanently accessed (downloaded, stored, printed, processed) toll-free by anyone Web-wide." [See also the definition at http://www.soros.org/openaccess/read.shtml.]

Open Access Colors

White: Journal that does not yet officially endorse author self-archiving

Green: Journal officially endorsing author self-archiving

Gold: Journal whose full contents are accessible online toll-free to all users.

[Source: http://www.openarchives.org]

OAI Open Archives Initiative

The Open Archives Initiative (OAI) is a protocol for collecting metadata about data files residing in separate archives. When the protocol is used by data services such as search engines, these sources can process the data in separate archives as if the data resided in just one archive. (In the technical jargon, the metadata harvesting protocol supports interoperability.) The BOAI supports OAI for all open-access literature, but BOAI is not part of OAI or vice versa. [Source: http://www.earlham.edu/~peters/fos/journals]

Open Source Software

http://www.opensource.org

"When programmers can read, redistribute, and modify the source code for a piece of software, the software evolves. People improve it, people adapt it, people fix bugs. And this can happen at a speed that, if one is used to the slow pace of conventional software development, seems astonishing." The Web site lists and links to many types of open source licenses. Open source officially means conformance to the Open Source Definition available at http://www.opensource.org/docs/definition.html

Other Sources of Definitions:

Peter Suber's The SPARC Open Access Newsletter

http://www.earlham.edu/~peters/fos/index.htm

Editorial Position of the FOS Newsletter / SPARC Open Access Newsletter, version 1.9 lists a number of definitions for major terms.

e-prints

http://www.eprints.org

Has its own lists of definitions: http://www.eprints.org/glossary/ and http://software.eprints.org/handbook/overview.php.

DOAJ Directory of Open Access Journals

http://www.doaj.org/articles/about#definitions

The aim of the Directory of Open Access Journals is to increase the visibility and ease of use of open access scientific and scholarly journals, thereby promoting their increased usage and impact. It has another set of definitions.

Budapest Open Access Initiative Frequently Asked Questions

http://www.earlham.edu/~peters/fos/boaifaq.htm#openaccess

The Million Book Project partners are not afraid of a good court battle, either. Taking on the United States copyright law in *Kahle v. Ashcroft* [http://cyberlaw.stanford.edu/about/cases/kahle_v_ashcroft.shtml], the project leaders proclaimed:

> Internet Archive, in partnership with Carnegie Mellon University, the National Science Foundation, and the governments of India and China, have been working on the "Million Book Project." The Million Book Project isn't focused on commercially successful books—those are available at bookstores. The project will include a number of books in the public domain—those that are free of copyright protection and thus usable without the need to obtain permission. But many books fall into a nether region. These are works that are not commercially viable and therefore not widely available to the public, but are nevertheless subject to continuing copyright protection. The Internet Archive wants to include many of these books, which we refer to as "orphan works," in the Million Book Project, but current law makes that very difficult [because] Works that have no continuing copyright value don't attract the interest of commercial publishers. They nonetheless remain subject to copyright-related burdens (i.e., the necessity of clearing rights) that prevent organizations like the Internet Archive from archiving them, preserving them, or making them widely accessible for study and creative re-use. Under our traditional regime of conditional copyright, these works would have been filtered out of the copyright system—

California Digital Library Services

The California Digital Library provides a wide range of services, such as CDL collections and services. These services include the following:

- ***CDL Image Demonstrator:*** A searchable and viewable collection of art and architecture images drawn from museum and art collections.
- ***Counting California:*** "One-stop shopping" for government data and statistics about California.
- ***Digital Preservation Program:*** The CDL hosts the UC libraries' digital preservation program to ensure long-term availability and access to digital content.
- ***Directory of CDL-Licensed Content:*** Lists the CDL's extensive collections of digital resources, including article databases, electronic journals, and reference texts.
- ***eScholarship Editions:*** Provides access to digital texts and monographs, including more than 1,400 UC Press titles. A significant number of the electronic books are available for free to the public.
- ***eScholarship Publications:*** Showcases publications from the eScholarship program, including interactive publications from the Electronic Cultural Atlas Initiative, legacy online journals such as the Dermatology Online Journal, and monographs from the UC International and Area Studies Digital Collection.
- ***eScholarship Repository:*** This free, open access repository infrastructure supports the full range of scholarly output, from pre-publication materials to journals and peer-reviewed series, by offering UC departments direct control of publishing.
- ***Melvyl Catalog:*** A searchable catalog of library materials from the 10 UC campuses, the California State Library, the California Academy of Sciences, the California Historical Society, the Center for Research Libraries, the Giannini Foundation of Agricultural Economics Library, the Graduate Theological Union, the Hastings College of the Law Library, and the Lawrence Berkeley National Laboratory Library.
- ***Online Archive of California (OAC):*** Federates collections of digital materials (such as manuscripts, photographs, and art) held in the libraries, museums, and archives across California through a searchable database.
- ***Request:*** Enables UC faculty, students, and staff to enter requests from the Melvyl Catalog and journal article databases for materials. Items not available at a user's home campus are delivered via interlibrary loan, while items available at a user's home campus are delivered through the document delivery service.
- ***SearchLight:*** Allows simultaneous searches across multiple journal databases, book catalogs, and other information sources available through the CDL and the UC campuses.
- ***UC-eLinks:*** Provides a way to easily move from an article or book citation in an article database to the full online content of the item, or, for print materials, to automatically look for a UC library location of the item.

many of these works would never have been registered, or, if registered, never renewed. But under today's unconditional system, there is no filtering mechanism to separate these works from commercially viable works that legitimately are the focus of copyright. So if the Internet Archive wants to include an orphan work in the Million Book Project, it must obtain permission from the work's owner. But figuring out who the owner is, and how to contact him, is difficult and expensive (especially in the absence of a reliable registry). Thus far, the difficulty of identifying rights-holders and clearing copyright under current copyright laws has largely limited the Million Book Project to government documents, old texts, and books from India and China, where copyright laws are less burdensome.

Because of this, Brewster Kahle, the chairman of the Internet Archive, and others have filed suit.[30]

Even nonprofit, academic, and library projects need revenue. The California Digital Library's eScholarship program states that it "seeks to develop a financially sustainable model...." Don't we all?

The ultimate question is not what kinds of projects libraries can come up with, but how, beyond grants and demonstration projects with strategic partners, the programs will survive past the initial funding stage. Remember NREN? That library- and academic-driven concept began with strategic academic, commercial, and government-funded demonstration projects, but failed to find consistent funding and so was left in the dust by the World Wide Web. What is the old saying? Lead, follow, or get out of the way. Leadership needs a solid funding base. Show me the money. After the stock offering, Google's got money. We're watching to see where it plans to spend it.

It Is Easier to Apologize than to Get Permission

Authors can always self-archive on their own site and/or their employer's site with pre-prints and, sometimes with permission from

publishers, for post-prints. But what if the research has already been published and nothing was said about self-archiving?

Professor Peter Jaszi notes, "In some cases, the fair-use doctrine in copyright law may actually not be adequate in its present form for the uses of researchers. ... The likelihood of litigation is low to begin with. ... There are no disciplinary rules of best practice for cultural historians or film scholars or medical historians. And, in the absence of that kind of collective understanding, it's no wonder that individuals give up the game before it begins."[31]

Fearing that authors will self-censure themselves before testing the fair use and self-publication boundaries, some encourage boldness. For instance, Stevan Harnad, in his listserv[32] message describing John Hopkins University Press' new contract giving authors self-archiving rights, noted similar thoughts in stretching permission to include posting to open access sites: "...specific objections can no doubt be negotiated by the author with JHUP for OAI archives like ArXiv or CogPrints, if the author wishes, either before self-archiving, or, more sensibly, after, if JHUP ever decides to ask the author to remove the article — as no publisher has ever done, for any of the 260,000 articles self-archived in ArXiv since 1991!"

Intellectual Access: The Other Issue

In his September 2004 "Cites & Insights," Walt Crawford makes the point that "open access" does not automatically mean the archived articles are intellectually available to a library's users.

My primary interest in this section is freeing up library funds so academic libraries can maintain humanities subscriptions, buy monographs, other books, and media, provide access to gray literature, maintain technical services and reference librarianship, and in other ways preserve the record of the civilization and maintain themselves as libraries. OA journals can help — *if* they're represented in library catalogs and *when* they replace overpriced commercial journals or force those journal publishers to reduce prices. As for OA archives, as far as I can tell, these are likely to have either *no* effect on library costs or — when they have one — a potentially disruptive effect on scholarly communication. As long as OA archives represent such a small percentage of the papers in a given subscription journal that libraries must retain their existing subscriptions, then the OA archives don't help the financial problem at all. When a large enough percentage of the papers in a given journal are represented in OA archives, and the OA archives are harvested so that libraries can reasonably expect to find those papers via OpenURL or otherwise, then a growing number of libraries can, will, and *must* cancel their subscriptions to those journals. The concept that libraries must *and will* retain expensive subscriptions as long as *any* significant papers are being published in those journals that are not available via other means is ludicrous in a world of limited library resources.

In an address to the Elsevier Digital Libraries Symposium in Philadelphia in January 2004, Deanna Marcum noted that the time has come to "build massive, comprehensive digital collections that scholars, students, and other researchers can use even more easily than they use the book-based collections we have built up over the centuries."[33]

The universal finding tools are just not here yet.

Conclusion

There are no solutions here. The real battlegrounds will be the courts,[34] Congress, federal agencies such as the National Institutes of Health [www.earlham.edu/~peters/fos/nihfaq.htm], protracted contract negotiations, and detailed "terms and conditions" for the end users. The litigation, lobbying, "strong arm" negotiating, and courting of public opinion will continue.

These are the problems of "living in interesting times" that the sage warned us about.

In truth, the players (authors, publishers, databases aggregators, open access archives, libraries, academic institutions, the Internet Archive, researchers, end users, etc.) have much more in common than many may think. To avoid blood on the walls from the war-games analogies that I have drawn, new alliances, partnerships, collaborations, and joint research and development projects should be explored. The players should "think outside their media," as LLMC has done in taking its microfiche collections to digital format through a partnership with the University of Michigan. While such a move may seem more evolutionary than revolutionary, the technological challenges and intimidating costs of such a move must have seemed daunting. With the Google digitization project now underway and promising to target all of the University of Michigan's book collection, other such projects may seem less frightening.

The potential changes in policies and increased open access to scholarly, scientific, technical, and other information could well spawn a revolution in research and education, the likes of which this world has never seen. First, second- and third-world countries could have universal access to a huge amount of high-quality, yet free, life-changing information.

Let the self-archiving authors, grant-funding philanthropies, professional associations, standard-setting-for-easy-searching-and-retrieval

organizations, gold and green publishers, and digitizing libraries lead the way. As for government funded research, watch for requirements that researchers post their results to free, open access Web information sites. And how about free full-text dissertations and master's theses written by the best and brightest joining the faculty publications and course materials? But jumping from one campus site to another in search of it is not what I call "access"! We need a more universal search-and-retrieval system than today's subject-specific and/or site-specific search systems. When I want to research the medical, scientific, and cognitive aspects of art therapy through color pixel-managed images of Russian icons, morphing by artificial intelligence developed through a new Linux-based open source operating system on the retina of infants, complete with an econometric model for developing and marketing the resulting technology, as well its potential impact on the Federal Reserve rate, I will need one-stop shopping. Is Google Scholar my future?

Today, interdisciplinary researchers must slog through numerous fee and free sites, use terminologies unique to each, and then still wonder whether they have retrieved all they could in each field of inquiry. The joint ventures, metadata tagging, data mining, with global and multidisciplinary links through and among the disparate collections of today, are the next step. The peer-reviewed, scholarly and scientific, self- and institutional-archived collections of full-text articles, dissertations, and faculty research and course materials are fruit ripe for the picking. Or, to maintain my battle metaphor, fields are awaiting plunder by the armies of end users searching for authoritative yet free knowledge.

When the dust settles, the former battlegrounds will hopefully be replete with verdant fields of medical breakthroughs, technical innovations, and progress in the arts and sciences, the likes of which have not been seen since the original Renaissance. ◆

In the end, this article represents only the first look at a war that promises to rage for years. In the April 2005 issue, you will find some reference material — a table listing "Some Major Players" and a "Keeping Up on Open Access" table that includes links to current awareness services, journals, blogs, and newsletters, as well as several of the background reports that I used to write this article and that I commend to your attention. It will also explore the history of a major open access support effort in the federal government.

Endnotes

1 For authors who want to self-archive, see "Self-Archiving FAQ for the Budapest Open Access Initiative (BOAI)" at http://www.eprints.org/self-faq/.

2 The United Kingdom's House of Commons, Science and Technology Committee's report, "Scientific publications: Free for All?," is a 118-page report on the state of scientific publication in England, following days of testimony. Volume 1 can be located at http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm, and in PDF format at http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399.pdf. Volume 2, consists of oral testimony and written evidence and appears at http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/399we01.htm

3 The Scottish Science Information Strategy Working Group's draft declaration on open access appears at http://scurl.ac.uk/WG/SSISWGOA/declaration.htm.

4 Under its mandate, the California Digital Library (the 11th library in the University of California System) negotiates "to enhance resource sharing of periodical literature within UC through systemwide access to all titles held at any UC campus ... through systemwide negotiation and licensing..." [http://www.cdlib.org/news/barriers.html]. See also New England Law Library Consortium [NELLCO] Web site, which states, "NELLCO's goal is to negotiate the best possible prices and access terms on electronic resources for the membership. ...Publishers and vendors benefit from consortium participation in a number of ways. First, publishers have a single point of access to many law libraries across the country. Your content will receive widespread exposure in a very short period of time. You can establish a trial with multiple law libraries through a single interaction. Second, the licensing terms for NELLCO members are negotiated through the consortium. A single license agreement is established for all subscribing members, saving enormous time and effort. Third, billing is centralized for NELLCO subscriptions through the consortium. ..." For other benefits and more information, see http://www.nellco.org/index.cfm?page=benefits.

5 Go to http://www.publishers.org/psp/index.cfm and http://www.pspcentral.org.

6 For the text of the bill, go to http://Thomas.loc.gov/ and search for HR 2613. Representative Sabo's introduction of the bill was short and well-stated, "Mr. Speaker, today I will introduce the Public Access to Science Act, PASA, of 2003, legislation to make federally funded research available to the public. It is wrong when a breast cancer patient cannot access federally funded research paid for by her hard-earned taxes. It is wrong when a family whose child has a rare disease must pay again for access to research their tax dollars already paid for. Common sense dictates we provide the most cutting-edge research to all who may benefit from it, especially when they have already paid for it with their tax dollars. The United States government funds basic research with the intention and the belief that the new ideas and discoveries that result will improve the lives and welfare of the people of the United States and around the world. Our government spends $45 billion a year to support scientific and medical research whose product is new knowledge for the public benefit. We must remember that government-funded research belongs to, and should be readily available to, every person in the United States. Lifting restrictions that prevent the widespread sharing of federally funded research can only speed scientific advancement. I urge you to join me by cosponsoring this legislation to require research substantially funded by the federal government to be ineligible for copyright protection, and thus available in the public domain."

7 Go to http://www.salon.com/tech/feature/2003/07/01/plos/index_np.html for the article (not free).

8 For an uncorrected transcript of their testimony March 1, 2004, visit http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/uc399-i/uc39902.htm.

9 No less than five standards exist for open access digitization projects. Access them free at http://www.soros.org/openaccess/software/. OSI, Guide to Institutional Repository Software. The guide describes the five open source, OAI-compliant systems currently available. As many institutions are developing repositories, OSI thought it would be helpful to produce such a guide so that each institution could select the software best suited to meet its needs.

[10] See "Top Medical and Scientific Societies Commit to Providing Free Access to Medical and Scientific Research" announcing the Washington KC Principles for Free Access to Science at http://www.dcprinciples.org.

[11] See Gale's primary materials, including the Making of Modern Law and Virtual Reference Library, at http://www.gale.com/

[12] Barbara Quint, "Google and Research Libraries Launch Massive Digitization Project," December 20, 2004, http://www.infotoday.com/newsbreaks/nb041220-2.shtml; "Google's Library Project: Questions, Questions, Questions," December 27, 2004, http://www.infotoday.com/newsbreaks/nb041227-2.shtml.

[13] Visit http://www.crossref.org/crossrefsearch.html and http://www.crossref.org/01company/09press_releases.html to see the various announcements.

[14] For the announcement, see http://www.umich.edu/news/index.html?Releases/2004/Mar04/r031004.

[15] For a detailed list of various funding sources and their projects, see Robert Martin's speech at http://www.clir.org/pubs/reports/pub119/martin.html.

[16] See http://www.infotoday.com/newsbreaks/wnd040823.shtml.

[17] See the eScholarship program of the California Digital Library [http://repositories.cdlib.org/escholarship/] and the collection of member scholarship programs [http://lsr.nellco.org/].

[18] Developed jointly by MIT Libraries and Hewlett-Packard (HP), DSpace is now freely available to research institutions worldwide as an open source system that can be customized and extended. http://www.dspace.org/

[19] John Markoff, "The Coming Search Wars," *New York Times*, February 1, 2004.

[20] Read the opinion upholding the jury's $19,725,270.00 award [http://www.mdd.uscourts.gov/Opinions152/Opinions/01-3898.memorandum.pdf).

[21] See http://www.mdd.uscourts.gov/Opinions152/Opinions/Lowrys_op0703.pdf for all the facts about photocopying, e-mailing, intranet posting, etc.

[22] For a sincere copyright disclaimer from a database aggregator, see Ingenta's at http://www.ingenta.com/popup_help/copyright_statement.html.

[23] The *New York Times v. Tasini* Supreme Court opinion can be found at http://supct.law.cornell.edu/supct/html/00-201.ZS.html

[24] The Endocrine Society has a special copyright policy:

> The Endocrine Society charges copyright permission fees for the use of copyrighted materials when the material will be used in a way that promotes or benefits a for-profit organization or when the dissemination of the material conflicts with the Society's ability to distribute it. The Endocrine Society does not charge: original authors to reproduce their work; individuals publishing with non-profit organizations; students. The Endocrine Society will review the information provided by the requestor for accuracy and completion. If the Society believes the provided information is inaccurate or incomplete, permission is not granted until clarification [https://www2.endosociety.org/apps/CopyrightPermissionRequest/index.cfm?].

[25] For ACM's copyright permissions, go to http://portal.acm.org/info/usage.cfm. You may also note, "As a matter of professional courtesy, the lead author of any ACM copyrighted work is consulted in weighing requests from third parties for permission to republish" [http://www.acm.org/pubs/copyright_policy/#Notice]. Nice touch.

[26] Go to http://www.crossref.org/crossrefsearch.html for information on Google's project with such organizations as ACM, BioMed Central, Blackwell Publishing, Institute of Physics Publishing, Oxford University Press, University of California Press, University of Chicago Press and John Wiley & Sons.

[27] Yahoo! and OAIster have announced a joint program. CAP enables Yahoo! Search to expand the breadth and depth of content users can access. Many of the scholarly collections included in OAIster were not previously indexed in popular Web search services and remained hidden from those who need the resources for their research. By enabling access through Yahoo!'s CAP program, these materials will be widely available to an international audience of scholars, students, researchers and enthusiasts. OAIster provides a direct link to an actual digital object — an image, book, document — not just a catalog or descriptive information. Examples of some of the collections currently available through OAIster include the arXiv.org Eprint Archive (an archive of physics research); Carnegie Mellon University Informedia Public Domain Video Archive; Ethnologue: Languages of the World; Library of Congress American Memory Project; and Caltech Earthquake Engineering Research Laboratory Technical Reports. In addition to the OAIster project, other participants in CAP include National Public Radio, Northwestern University, The New York Public Library, Project Gutenberg, UCLA, and the National Science Digital Library [http://www.umich.edu/news/index.html?Releases/2004/Mar04/r031004].

[28] While on the AALLnet site, check out the excellent article by Scott Burnham, "Copyright in Library-Held Materials: A Decision Tree for Librarians" [http://www.aallnet.org/products/pub_llj_v96n03.asp].

[29] See "Global Cooperation for Global Access: The Million Book Project" by Denise Troll Covey, associate dean, Carnegie Mellon University Libraries, at http://wwwoud.eurocris.org/conferences/cris2004/pdf/Cris2004-Troll.pdf.

[30] See *Kahle v. Ashcroft Case Page* at the Stanford Law School Center for Internet and Society [http://cyberlaw.stanford.edu/about/cases/kahle_v_ashcroft.shtml].

[31] See *The New York Times* article, "Permissions on Digital Media Drive Scholars to Lawbooks," by Tom Zeller, Jr. at http://www.nytimes.com/2004/06/14/business/media/14fair.html?exex=1088271146&ei=1&en=12f923c1e0963a0f.

[32] See the June 16 entry at amsci-forum.amsci.org/archives/American-Scientist-Open-Access-Forum.html.

[33] See Robert S. Martin's account at http://www.clir.org/pubs/reports/pub119/martin.html or the full text at http://www.clir.org/pubs/archives/dbm_elsevier2003.html.

[34] The most recent case of note is the $300 million lawsuit by Eliot Spitzer, the Attorney General of New York who sued GlaxoSmithKline for fraud in failing to disclose negative clinical trial information about the effects of Paxil on children. A Glaxo spokesperson said "the company would post data by Dec. 31, 2005, on clinical trials since the firm merged in December 2000. Test results from this week forward will be posted within 10 months of a drug's approval." The result is that other drug companies will now be posting the results of their clinical trials on their Web sites as well! See http://www.latimes.com/business/la-fi-glaxo27aug27,1,790959.story. On Aug. 3 /PRNewswire-FirstCall/ ran this:

> Eli Lilly and Company (NYSE: LLY) announced today that it would disclose the results of all clinical trials for which Lilly is a sponsor via a publicly available registry, beginning in the fourth quarter of this year. Consistent with the company's policy of open disclosure, the registry will include results of all Phase I through Phase IV clinical trials of Lilly's marketed products conducted anywhere in the world. Additionally, the company will begin posting the initiation of all Phase III and Phase IV clinical trials via the registry.
>
> "Lilly understands that patients, customers, and critics are looking for transparent answers that provide value to the healthcare decision-making process," said Sidney Taurel, Lilly's chairman, president and chief executive officer. "Our announcement today represents a comprehensive effort to publicly disclose Lilly's clinical trial information. These actions should prove to be invaluable for patients and the medical community as they seek to make informed decisions about Lilly medicines."

COPYRIGHT INFORMATION

TITLE: Open Access: The Battle for Universal, Free Knowledge
SOURCE: Searcher 13 no3 Mr 2005
WN: 0506004926002

The magazine publisher is the copyright holder of this article and it is reproduced with permission. Further reproduction of this article in violation of the copyright is prohibited. To contact the publisher: http://www.infotoday.com

Copyright 1982-2005 The H.W. Wilson Company. All rights reserved.