



A BIMONTHLY REPORT ON RESEARCH LIBRARY ISSUES AND ACTIONS FROM ARL, CNI, AND SPARC

## EXPANDING THE PUBLIC DOMAIN

by James Boyle, William Neal Reynolds Professor of Law and Faculty Co-Director of the Center for the Study of the Public Domain, Duke Law School

Remarks presented at the Association of Research Libraries 146th Membership Meeting, May 26, 2005.

During the last 10 years there has been a frenzied and intensive debate about the desirable limits of intellectual property policy. For much of that time, if you said amazingly bland and banal things like: we should have balance, or it's important to think about the inputs for creativity as well as protecting outputs, or we should not commoditize facts and ideas, you could be labeled as a communist, an anarchist, or, rather confusingly, both. So, what I think I'm going to do is produce a stunningly banal set of ideas. First, I'll discuss what we mean when we talk about the "public domain." Second, I'll explore a set of ideas about recent expansions in intellectual property policy. And third, I'll talk about public domain initiatives that we can undertake within private institutions.

### A Richer Understanding of the Public Domain

Although my topic is the public domain, I want to stress something that I would like you to remember throughout my talk: the public domain is fed by intellectual property. It is not merely the opposite of intellectual property. The way to have more things in the public domain is not always to get rid of intellectual property. For example, if we were to get rid of the patent system, many inventions would end up covered by trade secret law and we might never get access to them. Intellectual property has an important role, and that is the premise of everything that I'm going to say. Preserving the balance between intellectual property and the public domain is not an attack on intellectual property; rather, it's about preserving a living ecosystem between intellectual property and the public domain.

First, I want to pull back and address a few definitional issues, which I think need to be clarified in order to talk about the "public domain." We all have a rich and complex understanding of "property." We understand that there are lots of things you can do with property: giving it, sharing it. We know that we can rent an apartment, and that it's still owned by someone else, but we nevertheless have rights over it. We are, in fact, immersed in a culture of property, and it's constantly maintained, constantly named, constantly refined, all the way from "that's mine, you can't have it" on the playground through signing your first college lease to your mortgage and your retirement plan.

We also live in a world of the public domain—the realm of material that is not covered by intellectual property, and is accessible for all to use. But it is not as well named, and not as well understood. When we talk about the public domain, for example, are we talking only about complete works that are completely free, such as Shakespeare plays and Mozart symphonies? These are in the public domain in the sense that the copyright has expired, and you can do whatever you want with them. You can make a new version of them, abridge them, base a new work on them. We could also be talking about things which are not, and never could have been, the subject of intellectual property, such as  $E=mc^2$  or two times two equals four. Some people would include both the works of Shakespeare and Mozart, and the world of ideas and facts, in the public domain. Others might include the limitations and privileges within intellectual property as part of the public domain.

So, for example, my ability to criticize a book would mean that, for that particular use, it is in the public domain, or my ability to parody a song would mean the parody-able aspect of a song is in the public domain.

You might say, well, “what’s in a word?” The point is that we need to develop as rich, complex, and varied a notion about the “public domain” as we have about “property.” When we talk about the public domain, we need to ask, what is it that I want here? Is my claim that this whole thing should never be subject to rights? That this thing could legitimately be subject to rights, but at some point they should actually expire? That a particular use of an aspect of the thing should not be controllable? Often these ideas get conflated. It’s not that we need a precise definition of what is in the public domain. Instead, we need a better analytic process of definition. We should ask, what is our purpose here? What is the mental work we are trying to get done? What definition will get us there? And then make clear the definition we have adopted, and the tasks it seeks to accomplish.

We also need a richer understanding of the notions of the “public domain” as opposed to the “commons.” Until very recently, a lot of people would use these terms more or less interchangeably. But it’s not clear that they’re actually the same thing. Is open-source software in the public domain? No, not at all. It’s strongly protected by copyright—that’s, in fact, how open source can be maintained. It’s because of copyright that I can say, “The terms of this general public license are attached to your use of this software.” You may copy it freely, but if you wish to change it, you must add your new innovation to the ‘commons,’ not the public domain. You must make it available under this same license, which lets the future user, who also will add to the commons, use your innovation.”

Now the point is, that’s not the public domain. It focuses on many of the things that the library community cares about—access issues, sometimes price issues, sequential innovation issues—but it is built on the back of intellectual property rights.

In fact, there are currently developments in the scientific community, which some you may be aware of, where there is going to be a hard tactical choice along this front. For example, we’re right at the beginning of “synthetic biology”—creating entirely new molecules, entirely new biological entities, using, effectively, DNA as a programming code the way someone might use C++. Most of the sequences are probably not copyrightable. But some of the scientists who passionately want this stuff to be openly available wish that they were. Why? Because they want to attach a General Public License-like condition that says, if you want to use my building block, my enabling technology, then you have to add your innovation to the commons. They’re saying, this must be “property,” so it can be free.

So when we’re working on these types of issues, it’s important to be clear about definitions. And to be honest, I think right now we have a better understanding of the public domain than someone contemplating a society with property rules for the first time—say an anthropologist who’d come to us from outer space and had never heard of this weird idea of property. We have some familiarity with it. We’ve used it, we’ve been embedded in it. But we simply don’t have the richness and complexity, either of social uses, so that your kid would know what the public domain is all about, or even of philosophical, theoretical, and legal uses, so that we’d have a precise vocabulary and set of tools that would allow us to agree on particular definitions and goals, and get to work.

## Expansions of Intellectual Property Rights

To move on to my second area of focus, as you know, intellectual property rights have expanded dramatically in recent years. They’ve been expanding in every field of intellectual property, and in every dimension: length, extent of penalties, scope, subject matter. The copyright term has been extended by 20 years, and copyright penalties have become more severe. New rights protect not only the copyrighted work, but the digital fence in which the copyright owner wraps it. Patent law covers things we never used to cover—gene sequences, business methods. In the European Union, database protection now covers unoriginal compilations of facts.

What arguments have been used to justify this expansion? One is what I call the “Internet threat” argument, which assumes that, as copying becomes cheaper, intellectual property protection must increase. The argument goes like this. If you have a monk with a manuscript in his scriptorium, copying a book out by hand, you don’t need intellectual property protection, you just need to control a single copy of the manuscript. Copying would take months. Then along comes Gutenberg, and people can copy things quickly and more cheaply. We now have what economists call a public-goods problem, because we have a book that is non-rival and non-excludable. And now we see, for the first time, the need for intellectual property protection (which actually, somewhat confusingly, doesn’t arrive for over 200 years after Gutenberg). And as we go on, every time the copying costs fall, the need for intellectual property protection goes up. Zero intellectual property protection at the monk. The Statute of Anne by Gutenberg (except that it’s 200 years out of date), and as we go through the photocopier and the VCR, towards the world of Napster and Grokster, we need, effectively, perfect control. Because the Internet lowers the cost of copying to zero, and we have an infinitely leaky system.

Now, this is not a dumb argument, but it is wrong. It’s not dumb in that there is a real problem. The Internet

does lower the cost of copying, so it will magnify the amount of illicit copying. But it will also magnify the amount of licit copying. And it expands the size of the market, makes it easier for you to distribute things, lowers your advertising costs. On balance, are intellectual property holders better off or worse off? Well, even economists don't think that you can decide that in the abstract. They say you actually need evidence, right?

Here's another remarkable thing about intellectual property policy over the last 10 or 15 years: it is almost evidence-free. People criticize the FDA about Vioxx. But if we were doing FDA drug approvals the way we approved intellectual

property expansions, this is how the process would go. The drug company would say, "This is my friend. He took the pill and he feels better." Or sometimes even, "This is my friend, he needs to take a pill and he thinks it will make him better." And

then they would offer a model about as complicated as a picture of the person with a mouth and the pill in their stomach and say, "See?" That's about as data-intensive as things have been.

What if we had a test case where two regions adopted different intellectual-property policies, and we actually had evidence showing how these policies worked? Well, we actually do have such a case—in the area of database protection. In Europe, there is strong database protection under both copyrights and sui generis database rights. Many European governments also claim some kind of copyright over databases. And there is the idea that institutions, such as the Ordnance Survey or the weather companies, should recover their costs by charging users. The US tradition is totally different. In the US, there are no rights over data or unoriginal compilations of data. Any text produced by the government is free from copyright and passes immediately into the public domain. As for government-funded data, it is produced and distributed to the public with the idea, remarkably, that taxpayers have already paid for this, and shouldn't have to pay for it again.

Now, we actually have some good evidence about the effects of these different approaches. The United States database industry is considerably larger and more thriving, and has higher rates of return, than the European database industry. In fact, at the moment when Europe introduced sui generis database rights, there was a short one-time spike as database producers raced into the market, but then growth rates returned to previous levels, and many companies left the market.

And when did Reed Elsevier and Thomson enter the legal database market in the United States? It was after a case called *Feist*, which said that facts, and unoriginal compilations of facts, were uncopyrightable. That is to say, European companies chose to come into a classically public information field in the United States after they had found out, for sure, that they could get no copyright in unoriginal databases. Yet, even without database rights, they're getting high rates of return. So, we have evidence showing that less protection has been better for innovation than more protection. But you could spend days listening to arguments about database rights, and you'd never hear these facts mentioned.

Additional evidence shows that publicly generated data turns out to spur more economic activity if provided at marginal cost—close to zero—than if it is provided in order to recoup its cost of production. Europe puts into public weather-data

generation about half of what we do in the US, and it gets a nice return of about a six- to eightfold boost in production. The US puts in twice as much, and gets back a 39-fold increase in production. Why? The information is initially provided for free, but a massive secondary industry—the private weather industry—takes the publicly funded data and adds value to it. They employ more people, pay more taxes, and are an enormous portion of the economy. Keeping public information free just works better. It's not even a close call, as with Vioxx and aspirin.

So I have discussed two themes: first, the Internet threat argument, which says that as the cost of copying goes down, we automatically need more protection. And second, the idea that we can make intellectual property policy without having any evidence. This idea is bizarre: other government subsidies are rigorously assessed in order to figure out whether they're worth it, but here the government is handing out heaping slices of monopoly rent in the form of intellectual property rights, without empirical evidence that these rights are necessary, or that they will do more good than harm. My points are: lowering copying costs brings benefits, as well as costs. And we need evidence before we make policy. Banal and boring, right? It is in that context that I think we need to look at the range of intellectual property expansions that have been put forward, because in many cases we'll find that underlying them is the Internet threat assumption, and that they were passed without evidence. The call for evidence-based policy is one that we can really wrap our arms around: it's a positive proposal, and it's very hard to object to it.

*Preserving the balance between intellectual property and the public domain is not an attack on intellectual property; rather, it's about preserving a living ecosystem between intellectual property and the public domain.*

## Initiatives for the Public Domain

Let me turn now to private initiatives—practical things that we can do, akin to the kinds of things that the environmental movement did with its “think global, act local” initiatives.

### *Identification and Labeling*

One idea is to actually identify public domain materials as such, in order to make people aware that the public domain is there and that they are using it. We’re already digitizing things and making them available online but, for many people, the legal conditions under which they get this material are completely opaque. So we ought to tell people why and how they got access to this material, because if they realize that “this poem is here because the copyright term’s expired, and I’m glad about that because I can do something really useful with it”—then they can learn to value the public domain and what they’re getting out of it.

### *Fuelling Demand*

Along these lines, we also need to think more about the demand side of the public domain in general. We’ve thought a lot about the supply side—how to ensure availability and access. But what about the demand side? One of the things we found with Creative Commons is that an initial expenditure of time, effort, and money by people who cared could galvanize entire communities around public domain resources, or, in this case, resources made available under Creative Commons licenses. So, for example, we got David Byrne, the Beastie Boys, and other worthy musicians to put some of their music out under Creative Commons sampling licenses, which allow you to take snippets of a song, remix it, make your own song, and even, in some cases, sell it. Those musicians thought that would actually be great. So we ran a competition for all of the remixers out there who wanted to create something. They really got into it, and now there is a huge group of people, stretching well beyond those who were involved in the contest, who realize that there’s all this material out there that is free for them to use and to remix. We can stimulate the demand side of the public domain, and of the commons, by initiatives and educational approaches that get people to use public domain material.

### *Education*

At the university level, copyright education campaigns need to emphasize that copyright is a positive thing that people can use, rather than just something that gets in their way. These campaigns need to teach faculty and students about copyright, rather than merely telling them such untruths as there’s no such thing as fair use, and so forth. Copyright education is being done. It should be, it is important, but it is being done badly and inaccurately. It is being done in a way that is entirely foreign to the critical intellectual tradition of the university. Librarians

are ideally placed to develop strong, national, well designed, and visually attractive copyright education campaigns for schools and universities. Right now, copyright education consists of saying, “Don’t download songs illicitly, and if you do, turn off the upload feature.” That’s it? Sometimes the claim is that “no copyrighted material may be used without consent.” Really? What about fair use?

We need to be more serious about teaching our colleagues, our administrators, and our students about both sides of copyright. If we teach them that copyright serves valuable social goals, they might actually respect it more. If we explain the careful package of balances, the limitations, the boundaries of fair use, the things that can practically be done under existing laws, then we will begin to approach what copyright education should be all about. Copyright educators need to say, “These are the things you can do with the rights you have over your article, your materials. Here is what fair use allows. This is what you may not do, and this is what to think about. Here are the author agreements you sign. Do you want to? You have the right to self-archive. Are you doing it? Here’s this resource called DSpace, and so on.” Balanced copyright education is important.

### **Conclusion**

My goal here has been to offer a theory, and a practice, of the public domain. The theory and practice come with a change in attitude. It’s time to think about expanding the public domain, not just defending or salvaging it. Some of the decisions that have already been made were unfortunate. There was no need to extend the copyright terms, in my view. It was not economically justified, it didn’t harmonize the law, and we’ve locked up 20 years of culture for no good reason. But the good news is, I don’t think that the term extension would pass today. What we have to do now is to think of all of the ways in which we can use the wonderful technology that is available to us, and build a public domain that people can get access to practically, but also a public domain they are aware of. Because if people have a sense of this world of available, accessible information, and understand what they can do with it, not just as passive consumers, but as people who can actually use and build on it, then we will solve the theoretical problem I started out with. We will have our rich and complex idea of public domain because we will all be living it every day.

—Copyright 2005 James Boyle

This document is made available under the terms of a Creative Commons Attribution-NonCommercial-Sharealike License <http://creativecommons.org/licenses/by-nc-sa/2.5/>.

## CORNELL UNIVERSITY LIBRARY AND PRESS COLLABORATE ON "RACE, ETHNICITY & RELIGION" WEB SITE

by Kornelia Tancheva, Instructional Services Coordinator, Mann Library, Cornell University

Among the goals Jeffrey Lehman announced when he became Cornell's 11th president was a commitment to increase dialogue and understanding on campus about issues relating to race and religion. He urged students in particular to use their time at Cornell to deepen their understanding of these issues and cultivate the ability to respect and consider opposing viewpoints. To support the university's initiatives in this arena, in January 2004 the Cornell Library launched the pilot version of a Web portal (<http://racereligion.library.cornell.edu/>) as a resource for informed study and discussion of issues related to race, ethnicity, and religion.

The pilot site provided access to full-text books on race, which were published by Cornell University Press (CUP) and digitized by the library. Library staff also included suggestions for supplementary readings for students taking the spring 2004 courses "Judaism, Christianity, and Islam" (NES 251) and "Race in America and at Cornell" (GOVT 210). During the pilot phase, only students enrolled in those two courses were able to access the electronic CUP books.

After collecting feedback in the spring of 2004 from faculty members, students, and library staff, the project team released an expanded "Race, Ethnicity & Religion" site in September 2004. The new site includes seventeen full-text electronic versions of books on race-related subjects published by CUP from 1986 to 2003 and fourteen titles on religion issues—three of which are published by CUP. Among the digital books are *The American Dream in Black & White: The Clarence Thomas Hearings*; *Hispanas de Queens: Latino Panethnicity in a New York City Neighborhood*; *A History of God: The 4000-Year Quest of Judaism, Christianity, and Islam*; *I'm Not a Racist, but —: The Moral Quandary of Race*; *Skepticism, Belief, and the Modern: Maimonides to Nietzsche*; and *Fences and Neighbors: The Political Geography of Immigration Control*.

Cornell instructors will find a wealth of resources on race, ethnicity, and religion topics and can use the site as a pointer to supplementary readings for their students. The expanded portal includes resources on more religious and ethnic groups, as well as images from the library's collections. Since the site is devoted to issues at Cornell as well, students will find useful links to campus resources, departments, offices, and courses.

All Cornell students, faculty, and staff now have access to the e-books. Users can read the books online in HTML format or download the text as PDF files. The

portal also includes a full-text search option, which enables users to search not only the e-books, but also all the other resources on the site. For example, a search using the phrase "Hispanic American" produced results including books and journals in the library's collections, links to Web pages for Latino and multicultural organizations at Cornell, online reports based on the 2000 U.S. Census results, and e-books available through the "Race, Ethnicity & Religion" portal.

One of the major accomplishments of the "Race, Ethnicity & Religion" project in both its initial pilot stage and its second (and so far final) release was establishing and maintaining the organizational links required to make it a success. The project team consisted of library staff from different departments and with various expertise: subject selectors, Web designers, programmers, copyright clearance specialists, and managers. The team worked closely with the director of CUP to select and approve the titles to be digitized, as well as with a faculty advisory board, which included the professors teaching the two initial classes. The faculty advisory board members were appointed by the provost, and the project manager met with them in groups and individually to seek input and feedback throughout the process. Faculty were instrumental in selecting and recommending content, as well as advertising the collection to their students.

Professor Ross Brann, who teaches the course "Judaism, Christianity, and Islam," expressed his satisfaction with the way in which such collaborative projects can benefit the curriculum and instruction: "One of the challenges we face as instructors in assisting students with basic or advanced independent research is their tendency to head straight for the Web. Frequently they do so without the necessary tools to discern between what is valuable material and what is not. By contrast, Cornell Library's 'Race, Ethnicity & Religion' Web site offers the student a faculty-vetted, rich matrix of materials on a variety of interrelated topics."

The "Race, Ethnicity & Religion" project charts new territory in the library's ongoing collaboration with the press. Library staff are continuing to work with CUP and other publishers to identify additional resources that could be added to the Web site and hope to negotiate agreements that would enable users beyond the Cornell community to have open access to the electronic books.

—Copyright 2005 Cornell University Library

## PERMANENCE LEVELS AND THE ARCHIVES FOR NLM'S® PERMANENT WEB DOCUMENTS

by Margaret M. Byrnes, Head, Preservation and Collection  
Management Section, National Library of Medicine

*Editor's note: This is a slightly abridged version of an article that originally appeared in the NLM Technical Bulletin, no. 343 (March–April 2005), [http://www.nlm.nih.gov/pubs/techbull/ma05/ma05\\_archive.html](http://www.nlm.nih.gov/pubs/techbull/ma05/ma05_archive.html).*

The instability of resources on the Web is one of many challenging issues related to digital preservation. Several years ago, the National Library of Medicine (NLM) recognized the seriousness of this problem and included in its long range plan for 2000–2005 the following objective:

Take a leadership role in ensuring permanent access to important digital materials in health and biomedicine, including electronic journals, databases, documents published on the Web, and new kinds of scholarly communication and documentation of knowledge, using NLM's own electronic output and services as initial testbeds.

To this end, NLM has developed a system for communicating to users whether the resources they consult on the NLM Web site will be kept permanently available, change over time, or possibly disappear altogether. In addition, NLM has created an online archive for its permanent Web documents that are no longer current.

### Background

In 1999, the Working Group on Permanence of NLM's Electronic Information (a.k.a. Permanence Working Group) was appointed and asked to examine the range of electronic information produced by NLM and develop recommendations in the following areas:

- (a) Levels of permanence suitable for different categories of NLM information
- (b) Methods of recording and communicating the level of permanence of NLM electronic information
- (c) Procedures for ensuring that the levels of permanence are implemented in practice
- (d) Approaches to labeling, organizing, retrieving, and displaying NLM's electronic information so that the retention of older materials would not have a negative impact on those seeking current information

The Permanence Working Group's discussions focused initially on three important characteristics of Web documents: identifier validity, resource availability,

and content invariance. The group developed a rating system based on these three concepts. The ratings later were distilled into the following four permanence levels.

### Permanent: Unchanging Content

This resource will be kept available permanently. Its identifier will always provide access to the resource. Its content will not change. Example: Minutes of the NLM Board of Regents.

### Permanent: Stable Content

This resource will be kept available permanently. Its identifier will always provide access to the resource. Its content is subject only to minor corrections or additions. Example: Fact Sheets.

### Permanent: Dynamic Content

This resource will be kept available permanently. Its identifier will always provide access to the resource. Its content could be revised or replaced. Example: NLM's Home Page.

### Permanence Not Guaranteed

NLM has made no commitment to keep this resource available. It could become unavailable at any time. Its content and identifier could be changed. Example: Frequently Asked Questions.

The Permanence Working Group analyzed the documents that were available on the NLM Web site and developed a list of document categories. To simplify the assignment of permanence levels by library staff, document categories were assigned default ratings. For example, documents in the categories of announcements, news, applications, forms, calendars, and staff papers and presentations received a default rating of "Permanence Not Guaranteed," while such documents as bibliographies, databases, and digital library collections received a default rating of "Permanent: Dynamic Content."<sup>1</sup>

### NLM's Metadata Schema

During the deliberations of the Permanence Working Group, NLM's Task Group on Metadata and Methods of Recording Permanence Levels was appointed and charged with developing an expanded set of metadata to increase the retrievability of NLM's Web documents. It also was asked to decide how permanence metadata would be recorded and displayed. The task group recommended that metadata should be created for all publicly available electronic resources created by NLM and that permanence levels be a required element of the metadata set. The NLM set is based on the Dublin Core Metadata Element Set but with some local adaptations—most notably the addition of permanence ratings.<sup>2</sup>

### Implementing the System

A third committee, known as the Electronic Archive Group (EAG) then was charged with developing a pilot

project for assigning metadata including permanence levels and building an archive for outdated Web documents of permanent value to NLM. The EAG evaluated several systems under development elsewhere and concluded that TeamSite, a content management system developed by Interwoven, Inc., that was being purchased for NLM's main Web site, could be used for assigning metadata and managing the archiving workflow. A template was created in TeamSite and NLM Web contributors were trained to use it to assign basic metadata for all documents that would be submitted for promotion to the Web. The template is designed to minimize the burden on document creators. Default values or drop-down menus are provided wherever possible. When a contributor selects a document category for a document that has just been created or revised, the system automatically provides its default permanence rating. If a default rating does not seem appropriate for a particular document, it can be changed by the person responsible for assigning the metadata or by a system administrator.

When a contributor assigns to a document a rating of Permanent (Unchanging, Stable, or Dynamic content), the system notifies the NLM Archives Team. The Archives Team reviews the document category and permanence metadata and forwards the document for promotion to the Web. The Cataloging Section then creates a complete MARC bibliographic record with standardized access points, including Medical Subject Headings (MeSH) and an NLM classification number. The record appears in NLM's online catalog and is distributed to the bibliographic utilities and other NLM licensees. Enhanced metadata created by the Cataloging Section is then added to the header information of the online resource.

### **The Archiving Process**

The system prompts Web contributors at regular intervals to review and revise their current documents as needed. If contributors create a major revision of a permanent document or decide that a permanent document should be removed from the current site without being replaced, the archiving function is triggered.

When a document is moved to the Archives, the date archived is added to its URL. The only links in an archived document that continue to function are those to other parts of the same archived document. All other links are stripped when a document is moved to the Archives.

### **The Archives**

The Archives contain permanent resources with outdated or superseded content. This includes older material that was once on the current NLM site but is no

longer of current interest and earlier versions of current documents that have undergone major revisions. After investigating archives models developed elsewhere, the EAG determined that the best way to ensure proper migration of all permanent resources and allow searching and retrieval of archived items was to keep the Archives as a separate but integral part of NLM's main Web site. Archived pages are stored on a separate branch of the main NLM Web server.

The search engine was configured to query both the current site and the Archives but list the search results for archived documents separately. Clicking on an item in the search results takes the user directly to the archived document. Archives headers and footers indicate clearly to users that the documents they have accessed are no longer current. At the end of each document are publication, update, and archived dates as well as links to previous and more recent versions so that the user can trace changes in a document over time. Finally, if a user enters a URL for a document that has been moved to the Archives and there is no current version of the document on the main site, a redirect page will provide a link to the archived version.

### **Additional Work**

Currently only HTML documents are being archived. NLM has developed a sidecar approach to providing metadata for non-HTML documents such as PDFs. Contributors use a templated form similar to that used for HTML pages to enter metadata. System workflow validators require that contributors create this metadata file before a non-HTML document can be promoted. The metadata file is structured as Dublin Core XML schema, which can also be queried by the site search engine. Web documents created by the NLM administrative units that do not use the TeamSite content management system currently are not included in the Archives. In the future the workflow will be modified so that all of NLM's outdated Web publications of permanent value can be archived.

Finally, *NLM hopes to work with other libraries to encourage their use of permanence ratings for Web documents that are of lasting value.* For more information, contact Margaret Byrnes at byrnesm@mail.nlm.nih.gov.

<sup>1</sup> A table of the document categories and default permanence level ratings developed by the Permanence Working Group is available in the complete version of this article.

<sup>2</sup> See <http://www.nlm.nih.gov/tsd/cataloging/metafilenew.html>.

## SEEKING A GLOBAL PERSPECTIVE ON SCHOLARLY COMMUNICATION: CONTRIBUTIONS FROM THE UK

**H**ow do the University of Chicago Press's titles compare to Elsevier's in terms of median price? How long does it take first-time submitters to self-archive a work through the Internet? How do librarians and publishers feel about the concept of a national site license for a collection of journal titles? These questions about our current scholarly communication system are addressed in recent reports commissioned in the United Kingdom. It is worth taking a close look at three of these reports as much of the data collected and many of the findings are highly relevant for North American research institutions.

Oxford University Press commissioned a detailed study of the journal prices of 12 large scholarly publishers. Moving beyond traditional journal-pricing models, the Joint Information Systems Committee (JISC) sponsored two fascinating studies: one examining librarian and publisher perspectives on appropriate business models for journal content and the other analyzing faculty attitudes and behaviors relating to self-archiving and publishing in open access journals.

### Journal Prices in the Traditional Marketplace

Late in 2004, Oxford University Press (OUP) released the findings from a study of journal prices of 12 publishers of scholarly journals, both commercial and not-for-profit. Working for OUP, Sonya White and Claire Creaser analyzed eight major commercial publishers and four university press publishers.<sup>1</sup> While most of the data cover only a five-year period from 2000 through 2004, the pricing information is sliced and diced variously by broad subject area, by price per point of impact factor,<sup>2</sup> and by price per page. Each publisher's list was analyzed by quartile as well as by median journal price, and the top-priced journal for each was tracked.

When data are analyzed in that much detail some surprises are bound to emerge along with the broader picture. In general, conventional wisdom was affirmed by the observation that Elsevier had the highest median journal price for its list by a wide margin. Less intuitive was the documentation that showed Elsevier with the lowest rate of increase among the commercial publishers. In fact, two university presses posted higher rates of increase in median journal price.

Looking across subject categories, the commercial publishers generally demonstrated high median prices relative to the university presses. However, despite the obvious trend, universal truths are clearly rare as the University of Chicago Press had higher median prices than several commercial publishers in the arenas of biomedicine and science. Conventional wisdom was

again supported by the finding that the price per impact factor was generally substantially higher for the commercial publishers than for the university presses analyzed.

White and Creaser's quartile analyses of each publisher's title list are very unusual among pricing studies and provide a more detailed picture of pricing practices. Quartile analysis highlights the range of prices set across different titles in a publisher's list and tracks how price increases might vary between the most and least expensive journals on the list. For instance, most of the median price increase in a publisher's list could be the effect of increases in the most expensive titles alone. In theory, expensive journals might become less expensive while less-expensive titles grow more expensive over time. Despite the potential complexity of title-by-title pricing, the overall pattern shown by this study was overwhelmingly that publishers raise prices nearly consistently over the price range of their titles, i.e., all of a given publisher's titles tended to increase at about the same rate.

While the White and Creaser study holds few surprises for well-informed members of the library community, their substantial documentation along with some surprising details and unusual analyses make this work worth careful examination.

### Librarians' and Journal Publishers' Perceptions of New Business Models

It is perhaps beyond obvious that librarians and publishers have different opinions about the success and viability of possible new business models for journal publishing. The Rightscom study commissioned by JISC both documents the gap in perspective and looks at reactions to a set of potential new business models.<sup>3</sup> The business models considered range from a national site license to several pay-per-view options to several models that create open access.

The researchers conducted lengthy interviews with librarians from a wide range of higher education institutions. Similarly, interviews were conducted with journal publishers, both commercial publishers and not-for-profit publishers. These interviews yielded a varied list of observations, some generalizations, and many descriptions of diversity of opinion based on the type of institution represented.

It is no surprise to find that the librarians interviewed emphasized the need for wide access to a broad base of resources. Both pay-per-view, particularly user-based pay-per-view, and bundled models were not attractive to librarians. In contrast, publishers emphasized that declines in profitability were unacceptable and that greater overall levels of investment in journal collections were needed to accommodate growing volumes of scholarly output. Libraries and publishers tended to view each

other as excessively wedded to print publishing. Publishers reported they were neutral on open access.

One of the unique aspects of the study was the development of seven business models that were used to elicit reactions from librarians and publishers. Responses to the various models suggest the difficulty of building broad support for change. While some models seemed to offer few attractions to any of the respondents, none was broadly popular either. Even within the library community surveyed, significant variations were found in responses from different categories of institutions.

In general publishers and librarians alike objected to business models that impose constraints on usage and liked models offering predictability. Pay-per-view models were seen as problematic because of their tendency to constrain use and reduce predictability. Publishers were happy with bundled models and accepted consortial models, if not always enthusiastically.

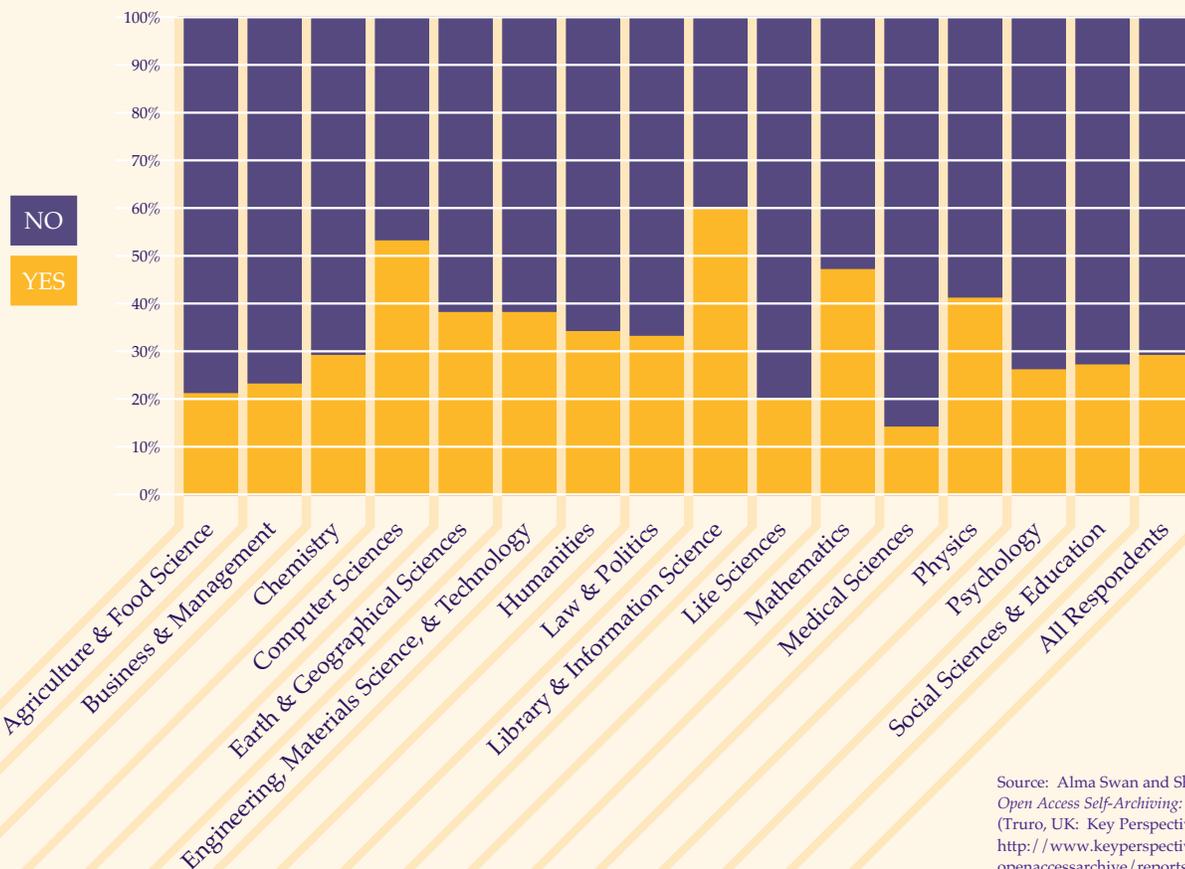
The report findings underscore that all business models involve trade-offs. Clear dissatisfaction with the

status quo was documented as well. Given the fundamental differences in objectives and concerns between publishers and librarians and the diversity of benefits obtained by different institutions within higher education, the findings highlight the complexity of identifying viable new models for journal publishing.

### Authors' Attitudes and Behaviors Regarding Self-Archiving and Publishing in Open Access Journals

Turning from the world of buying and selling journal subscriptions to look at authors, a fascinating study on author responses to open access was commissioned by JISC and reported by Key Perspectives, Ltd.<sup>4</sup> Using data collected late in 2004, the report is based on survey responses from almost 1,300 authors from around the globe. Only 7% of respondents indicated they were from the UK (27% were from North America). The study examined awareness of the ability to self-archive works and the attitudes and experiences of those authors who had archived works. Respondents also reported on their choice of open access journals to

## AWARENESS OF SELF-ARCHIVING AS A MEANS TO PROVIDING OPEN ACCESS



Source: Alma Swan and Sheridan Brown, *Open Access Self-Archiving: An Author Study* (Truro, UK: Key Perspectives, 2005), 45, <http://www.keyperspectives.co.uk/openaccessarchive/reports.html>.

publish articles. In addition, in several places the authors compared their findings to an earlier survey allowing them to report on trends over time.

Nearly half of the respondents reported having archived a work. Self-archiving was defined quite broadly to include both posting a work to a Web site and depositing a work in a repository that complies with the Open Archives Initiative (OAI). While Web posting was quite common, repository deposit was substantial. In many categories, deposit of refereed works had doubled since an earlier survey in January of 2004. These findings document that institutional and disciplinary repositories have made remarkable headway in changing scholars' behavior in a surprisingly short period of time.

Disciplinary variations were also tracked, showing an array of variations in deposit activity. Earth scientists, for instance, were most likely to have deposited a postprint in an institutional archive while medical scientists were most likely to have placed a postprint on a Web page.

Anxious to examine how onerous authors find self-archiving to be, the authors of the study gathered data on author perceptions of the ease of deposit and the amount of time required. Reassuringly, 54% of respondents described their first self-archiving experience as easy or very easy; however, 20% reported some level of difficulty. According to 75% of those who had deposited a work, it took less than an hour to archive their first work.

Since it seems that most authors have little actual difficulty depositing works, the question arises "Why don't more authors take advantage of self-deposit of their works?" The most common objective respondents cited for undertaking their publishing activities was to communicate their research results to their peers, an objective consonant with self-archiving. The answer to the question "Why not?" appears to be unawareness of the availability of self-archiving mechanisms. Of those who had not used self-archiving, 71% reported being unaware of the option. Lack of awareness of this option varied by discipline but ranged from 86% in the medical sciences to 40% in library and information sciences. The low level of awareness of self-archiving opportunities in the medical sciences is surprising in light of the announcements earlier this year by the US National Institutes of Health recommending public access deposit of funded research<sup>5</sup> and by the Wellcome Trust and the Research Councils UK mandating public deposit.<sup>6</sup>

Authors were also asked about their choice of open access journals as publishing venues. In the past three years, 24% of the respondents indicated they had published in an open access journal. The most common reasons for choosing to publish in open access journals were support for the principle, a perception of an

enlarged readership, shorter publishing timelines, and an expectation that citation rates would be enhanced.

Perhaps the most interesting question asked by the survey was how authors would respond to mandated deposit of works into OAI-compliant repositories instituted by employers or funding agencies. Nearly 80% of respondents indicated they would comply with such a mandate willingly while less than 7% indicated that they would not comply.

Overall, the survey paints a remarkable picture of the dissemination of the relatively new concept of author self-archiving. Uptake is happening quickly, with the main barrier being simple lack of awareness of the option. Authors who try self-archiving generally have a positive experience and tend to use the option again. Resistance to mandated self-archiving is very low among scholarly authors although unfamiliarity with the options is clearly a challenge. The findings reported here suggest that authors are likely to be supportive of mandates or recommendations for public deposit from funding agencies but there is substantial work to be done to increase awareness of archiving venues.

- <sup>1</sup> Sonya White and Claire Creaser, *Scholarly Journal Prices: Selected Trends and Comparisons*, LISU Occasional Paper no. 34, (Loughborough: LISU, 2004), <http://www.lboro.ac.uk/departments/dis/lisu/downloads/op34.pdf>.
- <sup>2</sup> Thomson ISI calculates the impact factor of a journal by dividing the number of current-year citations of articles published in that journal during the previous two years by the total number of articles published in that journal during the previous two years. For more information, see <http://www.isinet.com/essays/journalcitationreports/7.html/>.
- <sup>3</sup> Rightscom, Ltd., *Business Models for Journal Content: Final Report* (London: Rightscom, 2005), [http://www.nesli2.ac.uk/JBM\\_o\\_20050401Final\\_report\\_redacted\\_for\\_publication.pdf](http://www.nesli2.ac.uk/JBM_o_20050401Final_report_redacted_for_publication.pdf). Note: A presentation given by Hugh Look, of Rightscom, including some information not given in the report itself is available at [http://www.jisc.ac.uk/uploaded\\_documents/Hugh%20Look.ppt](http://www.jisc.ac.uk/uploaded_documents/Hugh%20Look.ppt).
- <sup>4</sup> Alma Swan and Sheridan Brown, *Open Access Self-Archiving: An Author Study* (Truro, UK: Key Perspectives, 2005), <http://www.keyperspectives.co.uk/openaccessarchive/reports.html>. Note: A set of summary charts and tables based on the survey data is available in a presentation offered by Alma Swan at <http://www.surf.nl/en/bijeekomsten/index6.php?oid=6>.
- <sup>5</sup> US National Institutes of Health, "Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research," February 2005, <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html>.
- <sup>6</sup> Research Councils UK, "RCUK Position Statement on Access to Research Outputs," June 2005, <http://www.rcuk.ac.uk/access/statement.pdf>; Wellcome Trust, "Wellcome Trust Position Statement in Support of Open and Unrestricted Access to Published Research," June 2005, [http://www.wellcome.ac.uk/doc\\_WTD002766.html](http://www.wellcome.ac.uk/doc_WTD002766.html).

# ARL ACTIVITIES

Kaylyn Hipps, *ARL Editorial & Research Associate*

## ARL TRANSITIONS

**Auburn:** Bonnie MacEwan was named Dean of Libraries, effective September 1. MacEwan is currently Dean of Collections and Scholarly Communications and Co-Director of Digital Scholarly Publishing at Pennsylvania State University Libraries.

## ARL/SPARC STAFF TRANSITIONS

**John D'Ignazio** resigned his position as SPARC Communications Specialist, effective August 16, to pursue a PhD in Information Science and Technology at Syracuse University's School of Information Studies.

## GOVERNANCE TRANSITIONS

**US National Commission on Libraries and Information Science (NCLIS):** On July 27, the White House nominated NCLIS Commissioner Sandra Ashworth of Idaho to a second term expiring July 19, 2009, and nominated Jan Cellucci of Massachusetts and Diane Rivers of Alabama to be members for terms expiring July 19, 2009. All three nominations require US Senate confirmation.

## OTHER TRANSITIONS

**The Andrew W. Mellon Foundation:** Don Michael Randel, President of the University of Chicago and a music historian, was named President of the Mellon Foundation, effective July 1, 2006. He will succeed William G. Bowen, who will continue his research and writing as well as supporting Ithaca Harbors, Inc., a nonprofit chaired by Bowen whose mission is to accelerate the productive uses of information technologies for the benefit of higher education around the world.

**National Association of State Universities and Land-Grant Colleges (NASULGC):** Peter McPherson, President Emeritus of Michigan State University, was named President of NASULGC, effective January 1, 2006. McPherson will succeed C. Peter Magrath, who is leaving NASULGC at the end of 2005 to become a Senior Adviser to the College Board and a consultant.

**US Federal Library and Information Center Committee (FLICC) & Federal Library and Information Network (FEDLINK):** Roberta I. Shaffer was named Executive Director. She was previously Director of External Relations and Program Development, College of Information Studies, University of Maryland at College Park.

## HONORS

**William Gosling**, former University Librarian at University of Michigan, was awarded the 2005 Library and Information Technology Association (LITA) Award for Outstanding Communication for Continuing Education in Library and Information Science.

## LIBRARY COPYRIGHT ALLIANCE FILES REPLY COMMENTS ON ORPHAN WORKS

by Prudence S. Adler, *Associate Executive Director, Federal Relations & Information Policy, ARL*

In May, 146 organizations, including the Library Copyright Alliance (LCA)—the American Association of Law Libraries (AALL), the American Library Association (ALA), the Association of Research Libraries (ARL), the Medical Library Association (MLA), and the Special Libraries Association (SLA)—filed reply comments in the US Copyright Office Notice of Inquiry on Orphan Works. The Copyright Office defines orphan works as those whose owners are difficult or even impossible to locate. Joining with 12 other organizations, LCA wrote in support of the Copyright Clearance Initiative that presents a framework for resolving the pressing orphan works problem (see <http://www.arl.org/info/frn/copy/orphanedworks/orphanreply.pdf>).

Approximately 650 of the 716 comments initially filed (91%) support the development of a legislative solution to address the orphan works problem. Importantly, there was a general consensus on the parameters of a solution to this issue among the diverse groups that filed comments. The joint filing stated, "Individual copyright owners and users, small not-for-profit organizations, and large commercial interests alike came forward with proposals that had remarkable similarities." The comments are available online at <http://www.arl.org/info/frn/copy/orphanedworks/LCAcomment0305.pdf>.

The Copyright Office conducted roundtable discussions on July 26–27 in Washington, DC, and August 2 in Berkeley, CA, to garner more public input on orphan works and possible legislative solutions. The Copyright Office sought input on four areas: identification of orphan works, consequences of an "Orphan Works" designation, reclaiming orphan works, and international issues. Robert Oakley, Director, Georgetown Law Library, and Jonathan Band, LCA legal counsel, represented LCA at the Washington, DC, roundtable. ARL President-Elect Brian E. C. Schottlaender, University Librarian, University of California, San Diego, and Gary Strong, University Librarian, University of California, Los Angeles, represented the University of California at the roundtable in Berkeley.

A report with recommendations on how to resolve issues surrounding orphan works will be completed by the Copyright Office by December 2005. Additional information on orphan works and the streaming presentation from the ARL, AALL, and MLA online conference on "Orphan Works: Issues and Legislative Strategies" held in May are available at <http://www.arl.org/info/frn/copy/orphanedworks/>. ARL continues to be actively engaged in efforts to resolve the orphan works problem.



*ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC* (US ISSN 1050-6098) is published six times a year by the Association of Research Libraries, 21 Dupont Circle, Washington, DC 20036. 202-296-2296 FAX 202-872-0884 <http://www.arl.org/newsltr/>  
Copyright: © 2005 by the Association of Research Libraries

*ARL policy is to grant blanket permission to reprint any article in the newsletter for educational use as long as the source, author, issue, and page numbers are acknowledged. Exceptions to this*

**Executive Director:** Duane E. Webster  
**Editor:** G. Jaia Barrett, Deputy Executive Director  
**Assistant Editor:** Kaylyn Hipps  
**Designer:** Kevin Osborn, Research & Design, Ltd., Arlington, VA  
**Subscriptions:** Members—\$25 per year for additional subscription; Nonmembers—\$50 per year plus shipping and handling.

*policy may be noted for certain articles. For commercial use, a reprint request should be sent to ARL Publications [pubs@arl.org](mailto:pubs@arl.org).*

## ARL CALENDAR 2005

<http://www.arl.org/arl/cal.html>

- |               |  |
|---------------|--|
| October 6–7   | The Future of Government Documents in ARL & Regional FDLP Libraries<br><i>Seattle, WA</i>            |
| October 25–27 | <b>ARL Board and Membership Meeting</b><br><i>Washington, DC</i>                                     |
| October 28    | Managing Digital Assets: Strategic Issues for Research Libraries<br><i>Washington, DC</i>            |
| November 4–5  | New Ways of Listening to Library Users: Tools for Measuring Service Quality<br><i>Washington, DC</i> |
| November 8–10 | Library Management Skills Institute I: The Manager<br><i>Los Angeles, CA</i>                         |
| December 5–6  | CNI Fall Task Force Meeting<br><i>Phoenix, AZ</i>  |

## ONLINE LYCEUM

Can't make it to our in-person events? Take a look at our Online Lyceum Web-based course offerings at <http://www.arl.org/training/lyceum.html>.

## ARL MEMBERSHIP MEETINGS 2006–2007

- |                                      |
|--------------------------------------|
| May 16–19, 2006, Ottawa, Ontario     |
| October 17–20, 2006, Washington, DC  |
| May 22–25, 2007, St. Louis, Missouri |
| October 16–19, 2007, Washington, DC  |

# 241

Cornell Library/Press Collaboration 5  
NLM Permanence Ratings for Web Documents 6  
Global Perspective on Scholarly Communication 8

August 2005