Overview -- Contracting for Content in a Digital World
Exploring the new issues for a new environment in policy and practice
Ann S. Okerson, Yale University Library

LITA Preconference Session on Licensing
American Library Association – New Orleans
Friday, June 23, 2006

As I reflect upon my charge this morning, to talk about the *current* state of licensing in libraries, I can't decide which of two tag lines that have been around since the sixties to use. "You've come a long way, baby" – which goes back to the first Virginia Slims ads in 1968 (when they were afraid at first to use the word "baby") – or "What a long strange trip it's been", introduced in the song Truckin' by the Grateful Dead at a concert at Fillmore West in 1970.

But this isn't the place or the time for history. Let us today mark that we've got a solid decade or more of experience with licensing. We have digested the idea that traditional print databases, journals, and now even books, have electronic versions; we've recognized that we can pay a complicated mixed fee for print, print plus electronic, or electronic only, and that there's no simple or single explanation for how such fees are set; we've embraced the idea that we can buy journals in bulk instead of subscribing by title; as well as the idea that librarians can gather together in consortia – to buy in bulk. I'm going to take for granted that everyone attending here knows nearly everything they ever wanted to know about the last ten years of that history and we are now ready to look around and to look ahead.

If ten years ago, we were novices, now we have a lot of good tools to work with. We've developed numerous model licenses, conducted seminars and workshops, and we are much more confident. Many things we worried about ten years ago have faded from the stage of our concerns. Everybody accepts the idea that we'll let walk-in users have access, that we're receiving and sharing usage statistics, that we're making aggressive consortial arrangements, and mostly we know how to deal with possible infringement concerns. When anxieties like that -- fears that used to loom large -- go away, we're all in a better place, and we should remember that.

So what are we worrying about today? I work on the front lines of these anxieties, watching the amount of investment we librarians make in digital resources climb and climb, balancing the needs of users with the realities of budget, and worrying that I might be taking things too much for granted. So, today I'm going to provide a short list of issues on my library's plate and also in our consortium – the NorthEast Research Libraries consortium, or NERL for short. Take this, please, as impressionistic and partial, knowing that I'm trying not to trespass very far into territories that other speakers will cover.

So here's my short list of topics:

1. **A continuing issue, not yet resolved: preservation and access**
2. **A looming issue, sometimes on the front burner: the future of the Big Deal**
3. **What to do when the tables turn: the library teaming up with the commercial publisher**

4. **Asimov comes to the library: what happens when my reader is a robot?**
5. **And for my closing sixties metaphor – the dawning of the Age of Aquarius, or, Open Access and its possibilities**

1. A continuing issue, not yet resolved: preservation and access

Perhaps this one has its own sixties tune, "Will you still need me, will you still please me, when I'm sixty-four?" The risks of loss of digital information are, as we all know, different from those of paper, and the fact that we as libraries do not *own* the digital information we license makes the anxiety worse. The magically powerful new content in our libraries might just go away completely, and we might not control it. One of the most important documents on this topic in the past year last year was prepared by Don Waters of the Andrew W. Mellon Foundation and a group of library directors dated 15 October 2005 and called "Urgent Action to Preserve Scholarly Electronic Journals." At a moment when more and more libraries are, holus-bolus, in the process of canceling print copies of journals (we're out of space, need to save money, and see the print copies piling up unread), getting to a more confident space is a more urgent issue than ever.

The licensing issue here evokes the terms "backup copy," "perpetual access," and "trusted third party". Different approaches are in play, with even a little partisanship emerging. The LOCKSS project out of Stanford (full disclosure: I'm on the LOCKSS board) distributes the task of ejournal preservation – or cacheing -- to many libraries, while the Portico system developed by Ithaka offers a service to libraries from a third-party vendor. A team led by Anne Kenney of Cornell will soon complete a major study of no fewer than ten significant e-journal preservation projects now under way, by way of giving us a grand environmental scan, and her paper, sponsored by CLIR, will be published later this summer. I commend it as thought-provoking and quite sensible. What the main systems have in common is that they depend on the publishers' cooperation, through a license of use, with the libraries as representatives of the reader community. This is a strange "space," though, where the licensing language and the cooperation of publishers and libraries runs well ahead of technological solutions, but we are making progress. At the moment – touch wood – this is an area where the interests of all parties seem if not perfectly aligned, then at least capable of reaching sufficient alignment to make us expect that we can get to a good place.

In technical terms, what this has to do with licensing is seen in our NERL standard license, available on our public Web site There we have reworded a key paragraph thus:

> Licensor will provide Licensee with one (1) copy in a mutually-acceptable format (or grant permission to make one copy) of any Licensed Materials holdings that are sold to another publisher/provider or discontinued for any reason, to fulfill Licensee's rights under Section XI, "Perpetual License", of this Agreement.
>
> In addition, upon termination of this Agreement or upon request, Licensor agrees to provide to Licensee in a mutually-acceptable format machine-readable copy of the Licensed Materials for Licensee's use. This copy shall be perpetually licensed to Licensee in accordance with Section XI, "Perpetual License", of this Agreement. Licensee is authorized to make such further copies in perpetuity as it may deem necessary for purposes of archival preservation, refreshing, or migration, including

> migration to other formats, so long as the purpose of such copying is solely for continued use and/or archival retention of the data and does not violate or extend the use rights contained in this Agreement.

And we then added a clause about trusted third-party archives service:

> Licensor acknowledges that Licensee may engage the services of third-party trusted archives and/or participate in collaborative archiving endeavors to fulfill the requirements of this provision. Licensee agrees to cooperate with such archiving entities and/or initiatives as reasonably necessary to make the Licensed Materials available for archiving purposes. Licensee may perpetually use the third-party trusted system to archive and restore the Licensed Materials, so long as Licensee's use is otherwise consistent with this Agreement. Licensor further acknowledges and agrees that, in using the third-party archival system, Licensed Materials may be made available to other system participants who indicate a right to those Licensed Materials.

2. The Big Deal – an offer we still can't refuse?

The promise of the Big Deal is this: pay what you've been paying all along for your individual subscriptions and get a whole lot more (or all the publisher's) content. The publisher's and library's initial advantage is the same: no haggling over individual subscriptions, just a single transaction. The caution that has emerged in the decade since we began negotiating these deals is, I sense, a suspicion that any thing that publishers are this eager to cooperate with must have something wrong with it – libraries *must* be over-paying. Certainly the Big Deal does not allow the kind of careful ratcheting down of traditional title by title subscription management. Where I used to know I had 200 journals from one publisher and could save money by dropping to 150, now I'm told that for the same price I get all 450 journals that the publisher offers – great! – but I have no choice at all and must continue to pay at the old rate. Some control has slipped from my hands.

At Yale just this spring, we engaged in an important exercise to explore this issue. In our case study, we examined closely Yale's relationship with a major international publisher, mostly but not entirely STM in content. The list price of buying all the publisher's titles, journal by journal, would have been about three quarters of a million dollar. If we added just the "Yale titles" – the ones that we chose to subscribe to in print, the price would have run to a little under half a million dollars. But in 2006, the total cost of our package deal was only a little over three hundred thousand, because of price caps negotiated over the last several contract terms. We came out ahead by about 150K. Usage continues to rise and cost per use continues to drop. We are able to provide relief for our "back room areas" such as cataloging (by taking all the MARC records through SerialsSolutions) and Interlibrary Loan-Document Delivery. Our users do not have to wait for content.

If this approach seems to validate, for now, certain Big Deals, for *our* particular library, at exactly the same time, we recognize that other institutions, looking at the same or similar publishers, may very well come to different conclusions. Yale is a well resourced research institution, and so our initial interest in this publisher began with a large subscription list and interest in a lot of journals.

We may be getting a good deal, but other, especially smaller, institutions might not feel the same. If or where I think the Big Deal works for our library, others may find that "by the drink" purchasing – guided by powerful search engines in a way that makes the whole publisher output into what is effectively a single database – works better for them.

3. What to do when the tables turn: the library teaming up with the commercial publisher.

This one's for Pogo – have we met the enemy, and is he indeed us? There are two pieces of history here, one obvious, one less so. The obvious part is libraries' long history as consumers, customers, people who pay retail and work to get a better deal for ourselves. In this history, it is easy to tell who the publishers are, whether they're the for-profits, the learned societies, or the university presses. That history is oddly comforting in that one might not always like the deal one got, but at least you knew who you were and where you stood.

Once in a while, libraries might have thought they were getting taken advantage of, though, in a sector of their dealings with publishers. I recall hearing the Librarian of Congress express frustration at deals done many years ago with LC's "content" (back before he was Librarian of Congress or anybody called it "content") -- when a major national commercial publisher with no academic pretensions at all literally set up a shop in the Washington area to produce commercial books for the mass market, drawing heavily on photos and documents from the Library's collections, reproduced for a pittance. The commercial publisher made a fortune, the Library picked up a nominal fee.

Times have changed. David Ferriero will speak later about the nuances of the relationship with Google that his and other libraries are building, and I won't trespass on that fascinating story, but I do want to suggest that it is part of a much larger story. Amazon.com, for example, is approaching libraries with the invitation that they become publishers. If we can identify materials that we want to digitize from our collections, with appropriate rights, Amazon will be the print-on-demand supplier of those to a retail audience through their website store. Recently both Cornell and the University of Michigan announced Amazon partnerships of this sort.

Once a library turns that corner we become ourselves part of the commercial publishing business, vendors of content and not just customers. Even as we still "give away" content to our defined user base, we are now also selling that information to *somebody*, and we have an undeniable financial interest in selling as much of it as we can, at as good a price as we can get. We may be noble people who settle for a little less than the most the market will bear, and some of us might even go back to charging nominal fees – only to find our provosts upset that we are not doing our fiduciary best for our universities.

Peter Kaufman of "Intelligent Television" prepared an important study of this emerging market trend in a paper called "Marketing Culture in the Digital Age: A Report on New Business Collaborations between Libraries, Museums, Archives and Commercial Companies". (http://www.intelligenttelevision.com/MarketingCultureinDigitalAge.pdf) In my view, these commercial relationships are going to be more and more an issue for many of us in the library – and now the word is appropriate – business.

So, if in 1997 the critical issue was how libraries would acquire licenses for material published by others, we see now emerging a reciprocal need, guidance for writing license contracts whereby others may have the use of library materials and property: contracts for libraries to use when publishers and others come to digitize the library's materials. We own some hot properties, no question. Not only does Google want to digitize: Gale, Readex and ProQuest are making deals left right and center, as are a number of smaller organizations. This is the 1970s reprint era on steroids. Right now I do not see that we as librarians have the expertise or the sense for the market that we need to navigate these issues, much less the consensus of our institutions about what is the right role for us to play. We need to work together soon on this issue to avoid either deeply unfair competition among ourselves or an unseemly scandal when we are seen to be taking money for services others think we should be providing for free.

4. You, Robot! Put down that book!

Let me take a step back for this one. One of the things we have negotiated for the last decade in licensing electronic resources is the way we count or identify or charge for our library's users. Some licenses allow only a finite number of simultaneous users of a given resource. Others set the price for unlimited access within an institution based on the number of Full-Time Equivalents – equivalent what? Well, equivalent somethings: faculty, faculty plus students, medical faculty and students only – or do we also count research staff, beanie babies, and the number of cars on the institution's parking lots on a weekend? I exaggerate only slightly, because somehow or other it is agreed that price should be a function of use or size or both.

We are now more comfortable with the way such negotiations work: we know how to have such conversations. But those conversations assume that our users will be doing with the information is relatively predictable: they'll read it. Click on a link, download to computer, print it – and read. Or skip the printing and read on screen. Usage of resources will go up somewhat because of accessibility – more people are now reading journals than ever before because they can read them at home on Saturday night with a wireless DSL modem while they lie in a hammock, perhaps, but there are only so many hours in a week, and there are better things to do in a hammock than read the *Journal of Neo-Baltic Paleobotany*.

What happens when a researcher decides to put a machine to work on the task of reading? In a recent conversation on the Liblicense-L list that I oversee, the visionary sometime publisher and keen-eyed denizen of Silicon Valley Joseph Esposito raised the question of licensing terms affecting data mining and text mining of journal materials. He wrote:

> My understanding is that materials (research data, user data, published articles, books, etc.) can be gathered together in such a way as to enable robots to sift through them and identify patterns and themes. These new patterns--effectively robot-generated discoveries--may include things that are not present in any single document in the collection. Thus, the collection is greater than the sum of its parts, but that greater value is only perceptible by machines. This past week I heard an excellent presentation by a biostatistician, who commented that human access to such databases is "of low value," in contrast to the "higher value of robot access."

To hear such an idea is to realize it has force and merit. One of the responses to Joe's message was the hearsay report that "the pharmaceutical industry is very interested in computational mining of the biomedical literature base." We can easily imagine how big pharma would set its robots to work searching the literature for connections that had escaped the naked eye or retentive brain of researchers, but that might lead to commercializable discoveries more quickly and effectively than depending on the old-fashioned stroke of inspiration.

And that's merely the direct text mining approach. Some are beginning to think about how to mine corpora for the things that aren't there. Will government agencies begin to assess the value of their investment in research by broad searches and analyses of the content of the scientific journals? Will bias and plagiarism and other misconduct be brought to light that way? We have no way of knowing, but when I heard this conversation, my first reaction was to say, yes, this will happen; and my second was to say, yes, Esposito is right, this will complicate the licensing.

Consider this: even at the mechanical level, if every academic institution in the country started sending its robots scouring through the databases of articles held online by vendors in their journals, the load on servers and network could quickly become appreciable. Now think about the big pharmaceutical company trawling for connections. Will the journal vendors think that selling their journals in the "traditional Big Deal" to all of us is a good deal for them when some of the "reading" isn't reading at all, but a new kind of sophisticated analysis? Will some publishers think they could do this work for us as another revenue generating source? Or at least think that they should charge by a different algorithm when they see the signs of that old phrase about "machine-readable text" being given a whole new and much more powerful meaning? I'll go back to the conversation on Liblicense-l and quote Cliff Lynch: "In fact, this kind of "reading" may represent the point of ultimate meltdown for copyright as we know it today."

Cornell's Phil Davis, another regular in our fascinating continuing seminar, observed that "Security software cannot distinguish the intent of data mining from stealing -- they both look like systematic downloading, and most publishers are pretty quick to stop this form of use." At the very least, such use would make nonsense of the "usage statistics" that we now depend on for assessing the usefulness of our electronic resource purchase. And Joe Esposito made one last point: is a robot's reading and utilization covered by "fair use"? Probably not..

5. The Age of Aquarius.

My friends know that I sometimes grow weary of the sectarian quarrels that brew and bubble on the Liblicense-l list over the topic of Open Access. But the issue of widest possible access is undoubtedly central to our debates and our future and even when the light generated is only a fraction of the heat, the topic deserves to continue to be explored and understood.

The experimental question, as I see it, is how we can make the fruits of scholarly and scientific research as freely and widely accessible as possible. In principle, few if any would oppose such a statement. But there is no agreement at all about the best ways to achieve such a goal, or even about the extent to which the goal can be achieved.

In institutional terms, we have two main choices: hold on to our product ourselves and publish those articles, books, data -- or conduct our negotiations with the journals our authors publish in and the ones we subscribe to a way designed to optimize access. If we hold on to the product ourselves, we are into the business of institutional repositories and we become publishers ourselves; if we go the other way, there are more possibilities. Let me propose a small taxonomy of four different models for improving the openness of access to scholarly products, and then say a few words about the implications of these models.

First, there is the self-archiving solution: the revolution in electronic publishing and dissemination as a do-it-yourself project. Taking heart from the example of Paul Ginsparg's marvelous work with the Arxiv(??) project, animated by the incessant drumbeat of Stevan Harnad's global advocacy, this solution holds that if scientists will archive their own work, measuring up to certain minimal standards, and assure the free distribution of it, the revolution will be achieved. All information will be instantly and globally available easily and for free. The main complications here are the concerns of institutional administrations for the costs and implications of institutional management of the repositories that are implicit here -- and also the sharp criticism of the technique by Harold Varmus, the Nobel Prize-winning scientist who gave Open Access some of its greatest early visibility and credibility when he was at the NIH and who still speaks relentlessly for it from his position as President of Memorial Sloan Kettering. Varmus does not recognize the articles deposited in institutional repositories as substitutes for quality journals, and it is these journals that he wishes to have openly available. Since Varmus's criticism became public, Harnad and others have been at pains to explain that Varmus is ill-informed: this does not appear to all readers as the most rhetorically effective position to take.

Second, there is the new business model solution: the creation of new journals, or the restructuring of old ones, with new business models that seek roughly comparable funding for journal publication, but reallocate the publishing costs. The commonest version of this new model is to ask authors to pay for publication – out of their grants or through their institutions – with various special deals, such as a reduced subscription rate to institutions whose authors publish with the journal or, alternately, reduced author fees for authors whose institution subscribes. The mixture of models here is confusing. The only observation to make is that if we keep the costs about he same but reallocate the paying, there will be winners and losers. I don't want to be one of the losers, nor do I expect any one else in the room here to volunteer for the losing side of this one.

Third, Big Brother solution: With what appears to be growing credibility, there is a movement afoot to resolve the issue of open access by governmental fiat, here and in Europe and other countries. On this model, nothing would change except that government agencies would require all articles initially supported by government funding to be made freely available, very likely on government-owned websites (but there are multiple models for that). There is no clear sense of how for profit and not-for-profit publishers would react to the change in traffic and demand that such a system might generate. As with the other models, it is difficult to predict the aggregate impact on the system of scholarly and scientific publishing as long as there are only experiments and not a wholesale conversion of systems. Cautious folk worry that moving precipitously towards a single governmental system of an insufficiently tested variety could be disastrous. I myself find it remarkable that scientists who bridle at government regulation and reporting requirements would voluntarily submit to another requirement, bringing in its train new forms of compliance

management, compliance officers, and the like. If the cost of such a government system, moreover, added to existing costs and reduced the funds available for direct support of research, scientists would not unreasonably object.

Fourth, and finally, what I might call the "From Each According to his Abilities" model – through a series of initiatives (some examples are OARE, HINARI, AGORA – you know them), a large community of publishers is working actively to make their journals available either for free or at very deeply discounted prices in emerging market and not-quite-emerging market countries. On this model, the costs of publishing are still be effectively born by first-world customers but considerably expanded access would be achieved. I mention these initiatives here, with full awareness that most true believers in Open Access would not recognize this as a part of their campaign. Nonetheless, I think a dispassionate observer, harking back to the principle of wanting to see access spread as widely and rapidly as possible, would perceive otherwise. These developing nations initiatives are undeniably putting more high-quality scientific material in more hands now, rapidly, than has yet been the case from any of the other segments of the movement I have described above.

In the Open Access initiatives, there are fairly immediate implications for licensing now, in the muddled year of 2006. As contracts are struck and money changes hands, facts are being created that will shape the possible futures we approach. We are already as libraries customers in a traditional way for some very untraditional products, like PLOS and BMC and other open access titles – still licenses, for all that their product and business model are different. If Open Access becomes a reality, will we still need licenses?

I don't know, but for the moment we can and should be assertive in writing into all of our contracts the most favorable authors' rights we can win. Consider language like this:

> AUTHORS' RIGHTS. Authors who submit scholarly content to publishers with whom we have licensed agreements should be permitted to retain copyrights to their work, to self-archive and/or deposit their works in institutional, subject-based, national, or other open repositories or archives (including the author's own web pages or departmental servers), and to use their work for other scholarly and educational purposes.

I think language of that kind is undeniably good in its own right; acceptable under current business models to publishers; and attractive in creating possibilities for future liberalization of access under any of the sectarian options we know or new options that may emerge.

Thank you.