# Cites & Insights

Crawford at Large

Libraries • Policy • Technology • Media

Sponsored by YBP Library Services

Volume 8, Number 11: November 2008 ISSN 1534-0937 Walt Crawford

Bibs & Blather

## Books and Blogs

Three items, some of which you may have seen on *Walt at random*. One of the items has been updated from the earlier post.

### Library Blog Books: Going OP

I may be a slow learner, but I can take a hint.

There have been no sales of Public *Library Blogs: 252 Examples* since June—and two sales of *Academic Library Blogs: 231 Examples* since June.

Barring a significant increase in sales, I'll accept that these books didn't serve a need and remove them from sale—probably around the beginning of 2009. That may make the (so far) 35 copies of *Academic Library Blogs* collector's items…or not.

I might be interested in doing a better, deeper, lateral version of the two studies as a single study—but that could only happen with sponsorship.

*Sponsorship opportuniies*

I'm still open to *The Liblog Landscape 2007-2008* having a sponsor, at the right price. Shoot me a line.

And I'd be delighted to work with the right agency on a combination of writing, data analysis/research and editing. Again, shoot me a line (waltcrawford at gmail.com).

### The Liblog Landscape: Moving Right Along

Here's what I said on September 24, 2008:

Phase 1 of the book is complete, resulting in a long chapter containing 607 brief blog profiles (to be revised later on) and one big Excel2007 workbook:

- A base sheet 607 rows deep and "S"–I guess that's 19–columns wide
- A derived sheet (references to some of those 19 columns, plus a bunch of derived numbers) also 607 rows deep but "AD"–I guess that's 30–columns wide
- A "working" sheet that starts each subphase as a values-plus-formats copy of the derived sheet, then gets manipulated to look at quintiles, build charts, create pivot charts, etc.
- A separate results workbook for results I know I'll need separately–because it's easier to have one workbook on one screen and the other on a second, in separate instances of Excel.

When I wrote that, I didn't realize I should also make a special spreadsheet for each graph that goes in the book, since Word wants to store the spreadsheet to support proper editing. (Yes, I lost one graph in the first five chapters because it was linked back to a spreadsheet that I then deleted. Ah well: Ten minutes to recreate the graph and save it properly.)

## Inside This Issue



So far? It's turning out to be quite interesting, but I'm not ready to start publishing intermediate results. Here's what I have so far in polished-draft form:

*1. Introduction and overview*

What this is all about, how I built the universe, some special notes (e.g., blog software used–and yes, there are still a few MovableType liblogs out there, although 18 doesn't compare to 230 WordPress and 222 Blogger liblogs), notes about authorship and affiliation, some graphs related to age of liblogs.

*2. How many posts?*

Looking at the number of posts in each blog in the two study quarters (March-May 2007 and 2008), including quintiles for each year, quintiles for the change from 2007 to 2008, graphs as appropriate, a list of blogs in the first quintile for 2008 and a few notes about "subsets" (e.g., blogs whose authors are affiliated with public libraries, blogs affiliated with medical institutions, etc.–I think there are six or seven subsets large enough to evaluate).

### *3. How long?*

A similar analysis–but this time there are two sets of metrics: Overall blog length and words per post, in some ways a more interesting figure.

Since September 24, I've also done:

### *4. Conversations*

Similar analysis of overall comments for a blog and the number of comments per post—and year-to-year changes in both metrics.

### *5. Getting the picture*

A brief chapter dealing with figures (illustrations, pictures, graphs) in each blog and figures per post—secondary metrics but still interesting.

### *6. Patterns of change*

This is the chapter I talked about (when it was Chapter 7) on September 24, as follows:

> I'm particularly looking forward to Chapter 7, a three-factor analysis that will actually show whether a significant proportion of liblogs have "fewer posts, longer posts, more comments per post" or whether that's a false or skewed supposition. (Well, it should show a lot more than that–but there are 45 different models for the three factors, even assuming only three values per factor: Significant growth, significant reduction, about the same. Yep, 45: 3×3x3=27, plus 3×3=9 (posts and post length) for blogs without comments and 3×3=9 (posts and comments) for blogs where length couldn't be measured. Fortunately, that should work out to three tables and a whole bunch of commentary…

I miscounted. There are **48** different patterns—including three where neither comments nor length could be measured.

The chapter now has two parts. The first takes a simplistic view (each factor up or down, with no "about the same"), portrays it in one table and discusses that table and its meaning. The second has four tables showing the 48 patterns and some of their combinations, discusses some of the meaning and lists blogs that fit into most of the patterns.

Since the first draft of that lengthy chapter is complete, I can now say my naïve hypothesis on liblog changes is correct. And that it's wrong. Well, you'll have to read the book…

### *7. Correlations*

This chapter looks at 45 correlations between pairs of metrics, using a straightforward test to find cases where there's a medium or strong correlation. For the seven cases where such a correlation exists, I provide the numbers, a scatterplot (cleared of extreme cases, with the modified correlation coefficient provided) and some comments about the correlation.

### *Still to come*

Chapter 8 looks at liblogs discussed in 2006 that are still around, considering changes from 2006 to 2008.

Chapter 9 discusses the "visibility" issue. No visibility or reach measures appear anywhere else in this book. I'll explain why, what's happened and consider a new minimal approach.

Chapter 10 may summarize conclusions and next steps, if there's enough there.

Chapter 11 (or several chapters beginning with 11) will include individual blog profiles, all of which exist but need updating.

### *Target?*

The book should be available from Lulu by the start of the ALA 2009 Midwinter Meeting—that is, January 23, 2009. I *hope* to have it available before then.

Yes, I will mount a list of the blogs somewhere on the web, either as an Excel spreadsheet or as a web page with links for each blog. No, barring sponsorship, I will *not* make the full metrics spreadsheet available for free. Even (most) OA advocates don't argue that people should carry out hundreds of hours of unpaid work for nothing, giving away the results because it's the right thing to do.

## Technorati and the Liblog Landscape

I've seen a number of comments on Technorati's recent *State of the Blogosphere / 2008* (technorati.com/blogging/state-of-the-blogosphere/). This year's report goes beyond most earlier ones (quarterly in some years, annual recently), with lots of analysis based on a survey of 1,000 bloggers. I'm going to ignore all that because it's not relevant to my own interests—that is, liblogs and library blogs.

Most observers seem to focus on polls demonstrating how "mainstream" blogs have become (which I don't doubt) and the growth in blogging–and ignore history, even though Technorati provides a direct link to the 2007 report and earlier reports.

Here, then, a few facts about blogs and related facts about liblogs. I assume Technorati's actual numbers are factual; I see no reason to assume otherwise.

### *Mostly a ghost town*

The "blogosphere" is much like Second Life: If you compare actual residents (active blogs) to counted residents (started blogs), it's mostly a ghost town.

What's "mostly"? 94% or more, depending on how you measure:

- The 2007 report said there were 70 million blogs as of April 2007, with 120,000 new ones emerging each day. If that 120,000 rate continued, there would be (or have been) about 120 million in June 2008, when the new study was done. The new study does not state the number of blogs or the number of new blogs each day–although it says "133 million blog records" since 2002, which presumably means 133 million blogs at some point. Technorati also quotes Universal McCann as saying that 184 million blogs have started as of March 2008. So let's say there are (or have been) somewhere between 133 and 180 million blogs.
- Meanwhile, Technorati says 7.4 million blogs had at least one new post within 120 days—a pretty modest measure of "active"—and just over 5 million posted in June. But if 120,000 new blogs were being created each day (each with at least one post), you could reduce that 5 million to a mere 1.4 million ongoing blogs. Of course, on that basis, 7.4 million is *smaller* than the number of new blogs during a 120-day period.
- Those figures make no sense, so let's be as charitable as possible and say between 5 and 7.4 million blogs are active, not just one-shot wonders. That's somewhere between 5.5% of 133 million and 3% of 180 million.

A liblog comparison? Excluding "friends and family" blogs, I come up with 533 liblogs (not library blogs) that were active (using a 90-day cutoff) in 2008—and at least 90% of those had at least one post a month. Have there been 9,700 English-language liblogs since 2001 (or 17,800 worst case)? It's possible—but I doubt it. Roughly 10% of the "visible" English-language liblogs that were active in 2007 had no posts during the 90-day study period in 2008.

### *40% drop in daily posts since 2007*

That's the truly impressive figure–and it does appear to be a direct comparison:

- In April 2007, Technorati counted an average of 1.5 million posts per day.
- In June 2008, Technorati counted an average of 900,000 posts per day: 40% fewer.
- Looking back, Technorati reported 1.2 million posts per day in April 2006–and 900,000 in August 2005.
- In August 2005, Technorati reported 14.2 million blogs and said 55% of them–or 7.8 million–were active. If that's right, the active blogosphere is basically as active as four years ago, with a lot of churn in between.
- That also comes out to about one post every eight days for the active blogosphere (although of course the average doesn't exist–only 1.5 million blogs had posts in a seven-day period)

Compare that with liblogs:

- For March-May 2007, 523 blogs had countable posts, for a total of 22,969 posts.
- For March-May 2008, 533 blogs had countable posts, for a total of 19,616 posts. (That 533 doesn't include 54 blogs with posts in 2007 but not 2008–and includes 64 new blogs and blogs without posts in the 2007 quarter).
- That's a drop—but a drop of 8.5%, which is a whole lot better than 40%!
- Those 533 blogs averaged about 213 posts per day as a whole, or about one post every 2.5 days per blog.

Overall, liblog posting declined at a much slower rate than blogs as a whole–and active liblogs are about three times as active as blogs as a whole.

### *Other tidbits*

- One-third of bloggers in general operate anonymously or with pseudonyms. That compares with 18% for liblogs.
- Roughly 85% of blogs as a whole have comment systems. Roughly 20% of liblogs didn't have any comments in 2008–but that includes blogs that don't allow them and blogs that just didn't have them.

### *What's the message?*

While liblogs have fewer posts now than a year ago—and for both liblogs and blogs in general, it appears that the peak was probably early 2007—liblogs are doing much better than blogs as a whole.

And when somebody blathers about hundreds of millions of blogs or says "everybody will blog in the future," feel free to ignore them. It's trivially easy to start a blog—but a lot of people (95%? 97%?) find, sooner or later, that they really don't have much to say that belongs in a blog. Why should that be a surprise?

## Making it Work

# Libraries and the Social Web

Set aside blogs and wikis for now. Those are established tools that can work very well for libraries, although it's not automatic. There are other "social" tools on the web, with libraries involved to various degrees. I'm catching up on some posts in these areas that I thought worth noting and commenting on—starting as long as 18 months ago.

### *Reference services and Twitter*

Stephen Francoeur posted this on May 2, 2007 at *Digital reference* (www.teachinglibrarian.org/weblog/).

> As a way of keeping in touch with friends, especially when you are out and about (such as at a conference, barhopping, clubgoing, etc.), the Twitter service is pretty nifty. But is there a way libraries might want to use this service to connect to users?

He notes that the Nebraska Library Commission is feeding its reference questions to a Twitter account (twitter.com/NLC_Reference)—and that's a little strange, since you get a stream of questions but no answers. (As of this writing, the account has 125 followers and has had 1,207 tweets since it began in March 2007.)

> Before discussing how Twitter might be used for reference services, I'd like to first note that our users might hesitate before adding the library as a Twitter friend if they intend to receive most of their Twitter messages via text messaging. As noted on the Twitter help pages, the Twitter service itself is free, but receiving the messages on your phone may incur charges depending on the plan you have with your carrier…
>
> Let's leave the issue behind of whether or not users would really want to add their local library as a Twitter friend; assume for the moment that they do. Is there a way that we could use Twitter to provide "reference" in some way? Perhaps…
>
> 1. User submits question to library…
>
> 2. Library answers the question and asks permission to add it (stripped of all personally identifiable information) to a publicly searchable knowledgebase…
>
> 3. With permission from the user, the question and answer are added to the knowledgebase, which in turn sends out its RSS feed, which itself is sent to rss2twitter to be passed along to the library's Twitter account.
>
> The Twitter account for the library would then be advertised on the library's home page. Library users who also happen to be on Twitter could add the library as a friend in Twitter so they can receive the stream of question/answer pairs…

Francoeur also suggests a second Twitter-only model—and privacy issues that could be involved. He was asking for reactions. The most useful comment was one from a librarian who keeps Twitter up at the reference desk and has gotten help from other librarians using it—but, in May 2007, she hadn't seen patrons using or asking about Twitter yet. I haven't seen a followup. Francoeur has a personal Twitter account.

Does this make sense? Accepting reference questions via Twitter isn't much different than accepting them via SMS or email (except that you can ask a much more detailed question via email). Responding via Twitter assumes pretty short answers and does, as Francoeur suggests, raise privacy issues. (One commenter, who liked the idea, didn't think there were privacy issues—"If patrons are willing to provide public questions in a Yahoo format, then they are not concerned about privacy.")

I haven't heard much more on this topic. A Google search for "twitter reference" (no quotes) in September 2008 yields the 2007 post as the first result. Looking at other results, I find a list of libraries with Twitter accounts—but those accounts are, with the exception already noted, used for library announcements and the like. (AskUsNow! in Maryland also has a Twitter feed with reference questions but not answers—a feed that appears to be open and might raise some mild issues, since it includes the usernames of people asking questions.)

### *do library users care about our new initiatives?*

Jessamyn West asked that question on July 18, 2007 at *librarian.net* (www.librarian.net/). She refers to a post that same day at *Tinfoil + raccoon* (rochellejustrochelle.typepad.com/copilot/) discussing the Wisconsin library survey I noted in the August 2008 *Cites & Insights*. As Rochelle Hartman noted:

> What is very interesting to me is that most users are not very interested in technology initiatives. You know, the stuff that many of us spend a great deal of time and energy using and promoting. The perceived value of these initiatives, however, is somewhat higher than actual interest.

Some of West's take (she raises other pertinent issues):

> One of the most interesting parts of the survey results is on page 16 entitled "New Initiatives" where they ask about how interested patrons are about using some new technology initiatives. To me they are asking all the wrong questions (mostly about content, less about context). They ask a lot of questions about downloadable content, which makes sense since the library probably has to shell out money for these things and wants to figure out if they're worth it. However, they also ask about 24/7 librarian access and IMing a librarian and also find that people tend towards the "slightly disinterested" side…
>
> Lastly, I'd like to point to the Internet question which was sort of glossed over. Of all the people surveyed 26% had no Internet at home and 23% only had dial-up. That's nearly half the respondents having a level of connectivity at home where a downloadable audiobook is worth basically nothing to them, and likely a group that doesn't spend a lot of time online. I'm not saying that we shouldn't still stress technology initiatives, but that's a pretty sobering takeaway when you're trying to provide more and more services online…

> As I've said before, I think that before we can fully immerse ourselves in a 2.0 initiative as librarians, we have to make sure we're counting the right things. If you only collect internal statistics on reference interactions that happen in-person or on the phone, it's no wonder that IM reference seems like a "flavor of the month" thing for the library to do. And, after the fact, if you can't show that people are really using the new techie things that you do provide it's harder to stress that those things that should be part of what your library is and does. Many of these things are countable—website stats, flickr photostream views, IM interactions—the question is: are we counting them?

I'll admit I found the first comment startling: A public librarian saying "The generation has been born that will not use books" and asserting that public libraries should emulate supposed "universities…that have no books, only data bases." As for West's comments in general, it's fair to say you need to count the right things—but that doesn't counter an apparent *disinterest* in new services.

Casey Bisson asked "Is it that they don't care? Or just don't want it from us?" on July 31, 2007 at *MaisonBisson* (maisonbisson.com/blog/). Excerpts:

> First, we've got to recognize some of the challenges of user perception that we face:
>
> - We've done so poorly leveraging current technology that our users can't imagine us offering novel services online.
> - So far, the new services we've offered—downloadable audiobooks and full text—have been presented so poorly, and our general web presence has been so weak that people just can't imagine using traditional libraries online.
> - Many of our library services are mediated… The web, on the other hand, is about freedom and self service. Our patrons may have trouble imagining a library delivering the things they value online.
>
> So what can we do?
>
> - Let's make sure our in-library services are outstanding…
> - Let's start small and make our library websites and catalogs as welcoming and easy to use as we want our physical services to be.
> - Then, perhaps, we can add a reference blog to help answer questions before they get asked. Perhaps we can offer some copy/paste code to allow users to embed books from our catalogs in their blogs or MySpace pages… Whatever we do, we've gotta make our basic online services great first, then we'll have the experience and knowledge (and user receptivity) to push new services online. And we'll grow and evolve our services from there.

I can't argue with Bisson's assertion that physical services should be outstanding as a first step. I'm less certain about "mediated services," since I think most of what most patrons use libraries for is materials and on-site database use (mostly materials), and with self-circ stations it's hard to call either of those mediated. Most of all, though, I'd agree that starting small and making sure you do it right is a good idea.

Jennifer Macauley also commented on West's post in "The technology our users want" on July 24, 2007 at *Life as I know it* (scruffynerf.wordpress.com/).

> Without doubt, trying to actually determine what our patrons want is quite a challenge. Trying to figure out if they care about technology initiatives, let alone 2.0 initiatives, is probably even more complicated.
>
> I've mentioned in the past that in my experience, patrons [at her college] (the majority of whom are 18-22 year olds) are not clamoring for 2.0 technologies--they are not pushing for the library to be on the "cutting edge."…
>
> Students primarily come into the library to use the computers--and to study… Sure, they use library resources in addition to their use of productivity software, etc. However, students actually use more non-library-related technology in the library than library-related technology… Currently, decent and knowledgeable technical support on all technologies that are available within the library seems to be much more important than adding dynamic content to our catalog or implementing virtual reference.
>
> My impression is that the majority of students could care less about Library 2.0 initiatives. Our experience mirrors that of Jeff Scott in that when asked about improving library services, the most common request is to extend the hours that the library is open. Extended hours, social spaces, cafes seem to be the things that draw today's students--not virtual reference, RSS feeds, blogs, wikis or interactive OPACs. Overall, this leads me to conclude that students aren't overly concerned about technology in the library. This makes it difficult for me to try and figure out how to plan for future technology initiatives…

One comment offered an interesting perspective: "I find myself thinking that it is not an issue whether our patrons want, or are clamoring for, the new technology so much as it is a matter of how we can use it to, hopefully fairly transparently, assist them in the 'ordinary' or more traditional library tasks, such as searching and retrieving both on-site and virtual materials."

Questions raised by the Wisconsin survey continue to resonate. It's certainly not universally reasonable to say, "They're not going to come, so we won't build it." But the Wisconsin survey supplies another, very different, instance of what's been clear in a number of cases: "Building it" is not enough and may or may not be the best use of available resources.

These are, to some extent, tougher questions than those involved in "web 2.0" services. Providing downloadable audiobooks costs real money, where blogs and IM reference (or Twitter reference) mostly just take staff time. But the questions need consideration in all cases—not to maintain a "culture of no" but to use resources in ways that best serve your own community.

### *Culture and social networking sites*

Chad Haefele posted this on January 17, 2008 at *Hidden peanuts* (www.hiddenpeanuts.com/). Excerpts:

> Some of the most interesting conversations I had at Midwinter were about the need for libraries to understand the culture of a social networking website before trying to market services through it…
>
> At my previous job, I experimented with creating a basic Facebook application that could search the catalog and a few other information silos. It worked great. Just one problem: nobody used it. Not one person installed the application who wasn't staff or a student worker.
>
> I didn't take the culture of Facebook into account when creating that app. Why do people go to the site? Not to do scholarly research, that's for sure. Or even if they did--why clutter up your profile with yet another application when it adds no value at all? Clicking to a similar search page on the library's website was a far more convenient process than navigating through Facebook's interface.
>
> So let's turn it around - instead of lamenting what students don't like on Facebook, focus on what they do like. One simple response is that they love widgets… They also love updating status messages, telling the world a little about where they are or what they're doing. So, why not combine these? Create an application which lets users pick a spot on the library floor plan and display that on their profile. "When I'm studying at the library you can find me here!" Students learn a little bit about the library's layout while finding their spot, and the presence of the widget reminds their friends that the library exists.
>
> Of course, I'd want to put a lot more thought into a project like this before going ahead with it. With social networking site endeavors, there is a fine line to walk. Some things are worth doing regardless of potential impact, just because they're so simple. For example: Set up a Facebook Page for your library with a Meebo IM widget… Even if only one single person ever takes advantage of the service, it is probably still a worthwhile use of time. But I ultimately threw away a lot of time on my Facebook catalog search widget, an end product which has had absolutely zero impact.
>
> We can't just wade into the middle of a social networking site and proclaim we know what is best for it…
>
> Different libraries can have very different user populations walking through their front doors, and libraries put a lot of effort into understanding those populations' needs. Why should the online world be any different?

Something in me wants to push back at setting up something that may only get one user because it's so simple to do—but I think the rest of the post is more important. Social networks and sites *do* have cultures, and it's not difficult to violate the norms of those cultures.

Kate Sheehan thought about these issues in a slightly earlier post at *Loose cannon librarian* (loosecannonlibrarian.net/), "are librarians culturally self-aware?" Brief excerpts:

> A big part of the 2.0 push is opening up the library, creating a two-way flow of communication and making the library about its users and communities, not its staff. But changing the cultural markers around libraries is a bigger project, especially when a lot of what we do offer is, well, workish.
>
> It's easy to become enamored of social networking sites and Web 2.0 toys to the point where they seem like a panacea for everything that's wrong with your library or your job. Slap a wiki on it and call me in the morning. The most successful uses of the newest tech tools have recognized that they're just that: tools. Midwinter has me revising that to add they're tools with their own cultures. Librarians are frequently hyper-aware of the internal culture of their own organizations, but are we less vigilant about our cultural significance outside of the library…and are we as mindful as we should be of the culture of the online communities we're trying to leverage to promote library services?

In comments, Michelle Boule provided a succinct summary of the change I'm seeing in librarian attitudes toward the whole 2.0 thing:

> …We have to have a service we want to provide and then choose the best tool. Before implementing the tool, we should understand how it works, the culture behind it, and how to maintain a level of control to keep it useful (reduce spam, etc.). I see so many libraries jumping into new tools, like MySpace, Wikis, or blogs, without considering any of these things and then when it fails, they think it was the technology not their lack of planning.

In a sense, it really *is* the technology that's gone wrong—because it's perceived as a more general solution than it really is. There are other things to consider before implementing some social-web tools, such as library commitment to make it work and whether there's a real need or something concocted to make a tool look like a solution, but on the whole Boule's right.

Meredith Farkas, no more a Luddite than Boule but also a long-time voice for thinking things through, added (in part):

> …While I'm a big booster for social software, I don't think these technologies are useful in every library, and I think we need to consider our very unique population (and sub-populations) when we consider any of these tools. I'm a big fan of libraries having student workers or teens staff their MySpace and Facebook presence, but honestly, I think patron level of comfort with librarians in social networking software differ library-by-library.

> I think we can frequently develop tunnel vision when working in libraries, and that's why it's so critical to **talk** to our patrons, to get out of the library, and to visit other libraries to remember what it's like to be a patron. We hear these stories about libraries implementing blogs and wikis and all this cool stuff, and we start to think that it's the blogs and wikis that are cool, not how they are used and what need they are fulfilling. It's not the tool that matters; it's what we do with it and how it meets patron needs.
>
> I'm glad these questions are being asked. We need to be critical of our role in the "2.0" world. We want to do right by our patrons, not force ourselves on them, invade their spaces, or create tools they don't want to use. And sometimes it's difficult to know if what we're doing is helping, hurting or is just plain getting ignored.

Part of the need is advance consideration—not huge planning efforts but a reasonable attempt to be sure there's a *real* need or desire and the tool reasonably addresses that need. Part is followthrough—a commitment to use the tool effectively. (A Facebook profile that's never updated or is obviously being updated only as a PR outlet is probably no better for a library than a blog that has four posts a year.) Then there's continuing assessment. What happens when what you're doing is "hurting or is just plain getting ignored"?

### *Is Facebook this generation's Rolling Stones?*

In some ways, this post—by Melissa Mallon at *ACRLog* (acrlblog.org/) on January 23, 2008—may be a bridge between two related sets of posts: The group just discussed and another group, right around the same time, referring to a University of Michigan study.

Mallon considers the extent to which teens are "growing up virtually" and continues:

> Since libraries started creating Facebook and MySpace pages, I have felt rather conflicted about the whole idea. I understand the theory of wanting to connect with students where they are (because, obviously, Facebook is where they spend most of their time), but I've always wondered if it can be truly effective. I'll be honest: I'm 25, I use Facebook to stay connected with friends, and if I were still in college, I wouldn't be "friending" my professors and librarians. Hearing the interviews on Frontline only confirmed my suspicions that teens and young adults don't want authorities online…
>
> …I am surprised at times to feel a disconnect with the students I'm teaching, when many of them are less than a decade younger than I am. When I think about it, though, I doubt many of the students would want to hang out with me in real life; why would they want to hang out with me virtually? Thus, I've come to this conclusion: instead of "joining" students in their virtual space, I think we should try to focus on catering to their virtual learning styles. Whether this means offering more online workshops, or virtual reference services, or blogs and podcasts, I'm not sure. I'm just not convinced that implanting our libraries into Facebook and MySpace is making quite the splash we think it's making. But, hey, I'm just a new librarian!

In comments, Joan Petit argues for librarians being on Facebook to better understand users—but "I agree that librarians shouldn't be out there friending patrons."

> But many librarians have set up place pages for their libraries, so students can become "fans" of the library. I've just set up a page like this at my library, and we already have a few dozen fans. That's a great way for students to get our hours, contact information, etc, and it helps them feel more connected to the physical library, even when they're online at home.
>
> Finally, there is also a role to play for librarians in helping our students understand privacy implications of social networks. Despite their technological sophistication, many otherwise bright students aren't doing enough to protect themselves online. I've shown students how to ramp up their privacy settings, which I know how to do because I follow these technologies myself.

Kelly A., a 24-year-old, says "even as a librarian I have zero interest in friending my director or any of the other faculty/staff from my university who have Facebook. Facebook is for friends--for social networking. Students don't want us there trying to be cool." She did create a library MySpace page "about a year ago" and put up notices near the computers—and got one friend.

> There is a lot to be gained from understanding how students are using these tools so we can better understand and cater to their learning styles and preferences. But that's very different from jumping into their turf and trying to get accepted.

Offering a different perspective, char booth noted survey results at Ohio University indicating student enthusiasm for library search/help applications in Facebook—which, to some extent, supports the argument for "catering to their virtual learning styles."

### *Social networking for libraries a bust?*

That's Sarah Houghton-Jan's title for a January 14, 2008 *LibrarianInBlack* (librarianinblack.typepad.com/) post based on a University of Michigan student survey. This continues a theme already stated—but in a different light.

> One question found that by and large the respondents use social networking sites, but the majority (76%) would not respond to a library presence on Facebook or MySpace, either because existing methods of contact were sufficient or because these tools are social networks and not places for library invaders… I would be very interested to see more surveys of this type, across organizations and locations, to see if this trend holds up...It's possible that we were wrong to believe that a social networking tool would attract all of its users to our services.

Some comments focused on the 24% (or 17%) who would, or might, contact a librarian on a social networking site. Houghton-Jan agreed that such presence may be desirable (unless the "unwanted authority figure" issue comes into play), but noted that 17% is much lower than the kinds of predictions that have been made. "While 17% is nothing to sneeze at, it is certainly below expectations…which tells me that the expectations were impossible and ridiculous to begin with." Anne-Marie looked for

> some conversations about why librarians might want to use these sites even if people don't want to talk to us this way. So some initial hypotheses were challenged - new hypotheses = new thinking.

Angel Rivera commented in "So, social networking and libraries may not be such a big deal," a January 18, 2008 post at *The gypsy librarian* (gypsylibrarian.blogspot.com/):

> The finding from the University of Michigan survey…should not come as a surprise. This is not a new finding; it simply validates what a few brave librarians who don't go "ga ga" over every 2.0 shiny gadget already knew…
>
> The people who won't give up will say that there is still about a fourth of users who would respond… The point is that 17% is not exactly great return on the investment (or ROI for those who like to use business jargon). I am not saying we neglect the 17%, but the evidence shows this is not a groundswelling wave of users needing or demanding our presence in places like Facebook…
>
> …This is only one survey. There are anecdotal accounts in the library sector of the blogosphere claiming some degree of success. Some of those places should conduct surveys of their own and get some actual evidence. Would that new evidence replicate the Michigan finding? I would like to know, but I'll say that, at the moment, the evidence is telling us that social networking is not the hot frontier for libraries it has been made out to be. It's nice, but it's not such a big deal.

Finally for this cluster, Simon Chamberlain offered "Social networking a bust?" at *Simon Chamberlain's library weblog* (chamberlain.net.nz/blog/) on January 18, 2008. He thinks the 17% figure is promising, and continues (in part):

> However, it does indicate that we may have set our sights too high in terms of the benefits of using these sites, and that 'going where our users are' may not be sufficient to get users using our services, if our users don't want us in that particular space. I also wonder if the 17% might mostly be students who would use the library anyway, whether they used the print collection or databases, websites, whatever.
>
> …Social networks are mainly used as a means for users to develop and display their identity… Users aren't necessarily using these sites to solve problems or find information, they are using them for social purposes. It makes sense, therefore, that many users might not find library applications interesting or relevant.
>
> And, of course, libraries are most emphatically not cool, which may just make some users reluctant to become our friends, or fans.
>
> That's not to say we shouldn't be there. The cost of maintaining a social networking profile would seem to be fairly low, and if it results in even a small benefit, it's probably worthwhile. My current thinking is that it might be most useful as a way for librarians (not libraries) to build relationships with students (and faculty?) in their subject areas. Most of the postgraduate class, and a number of undergraduates, know me by name or face. By "friending" them, I would give them an easy and unobtrusive way to keep updated with what was going on in the library (new databases, etc) - or even just to let them know that I was going to lunch (and save them a fruitless walk from the computer labs to the reference desk, looking for me).
>
> Of course, that's me speaking as a reference/liaison librarian, and I should probably remember that there's more to an academic library than just what I do. But from my perspective, at the moment, that approach seems most useful…

The lesson here is *not* "stay away from Facebook." I take away at least two possibilities:

- Don't expect miracles—particularly not if the library itself attempts a Facebook presence.
- At least in academia, it's possible that *librarians* building relationships with their own circle of users is more promising than *libraries* attempting to befriend the community.

### *Ten social networking tips for libraries*

Does it make sense for your library to have a Facebook or MySpace profile? Maybe, maybe not. On one hand, "the adults should stay away" as a Facebook issue is a little old hat (as it is for MySpace): Almost certainly most users of both services these days are adults. On the other, trying too hard to make friends can either be spam or just plain annoying.

If you're going to do it, do it right. To that end, this July 28, 2008 post by Sarah Houghton-Jan is a fine set of reminders. She offers ten paragraph; I'll just list the topic phrases:

- Do your research.
- Keep your information current.
- Use a photo of a real live person.
- Feel free to have an alias.
- It's not all about your friends.
- Suss it up (make sure you're notified as much as possible).
- Have fun!
- Know no fear.
- Deleting accounts is okay.

If those sound intriguing, go **read the whole post**. And do read the comments—including one pointing out that maybe a failed attempt should be "left dark" rather than deleted, so a spammer can't take your deleted name. I've seen former liblogs turn into adlink pages; it's probably better if, in the event that your library Facebook page turns out to be more trouble than it's worth, people aren't led to believe that your library is now marketing "size enhancement" products.

### *Not without the Web*

If this episode seems negative, it isn't intended that way—but maybe that's why much of this material sat around so long before I used it. (My personal Facebook profile? I'm still thinking about it…)

So, as a fine example of the virtues of the web, I point you to this July 28, 2008 post by Meredith Farkas at *Information wants to be free* (meredith.wolfwater.com/). She offers 800 words on what she probably wouldn't have done without the web and what it's meant to her. It's a fine story, augmented by a number of comments on what the web has done for others. **Well worth reading**.

## Perspective

# How Common is Common Language?

You know Google Book Search—and, realistically, Google itself (which includes book results)—has been touted as a fast, easy way to check for plagiarism. Just do a phrase search for a distinctive sentence or long phrase, and if you get a match you should check further for likely plagiarism.

Some writers seem to take it a little farther. Plug in a distinctive five-word phrase; if there's a match you've got a plagiarist. That's nonsense. Few five-word phrases are all that distinctive, and I'm not sure it's reasonable to call five borrowed words plagiarism.

Still, the general approach seems sound. With several million digitized books and billions of other text sources, and with phrase searching that appears able to accept fairly long phrases, it's a good first step (if only a first step).

## Distinctive Sentences

But what's distinctive? How do you identify sentences that are good candidates for checking? Let's turn that around: Aren't most sentences common enough that they're useless for detecting plagiarism?

I can't take credit for that idea—although, in one of those "accidental plagiarism" situations, I'd nearly forgotten the seed of it. Fortunately, Google comes to the rescue. Paul Collins wrote "Dead plagiarists society" on November 21, 2006 at *Slate* (www.slate.com/id/2153313). Collins focuses on Google Book Search, noting the extent to which it had already aided folks in uncovering published plagiarism. He offers some examples and suggests that we may see a lot of newly detected plagiarism in the future.

> But wait, you might ask, don't people accidentally repeat each other's sentences all the time? *It seems to me that this should not be unusual*. Yet try plugging that last sentence word by word into Google Book Search, and watch what happens.
>
> *It*: Rejected—too many hits to count
>
> *It seems*: 11,160,000 matches
>
> *It seems to*: 3,050,000
>
> *It seems to me*: 1,580,000
>
> *It seems to me that*: 844,000
>
> *It seems to me that this*: 29,700
>
> *It seems to me that this should*: 237
>
> *It seems to me that this should not*: 20
>
> *It seems to me that this should not be*: 9
>
> *It seems to me that this should not be unusual*: 0
>
> *It seems to me that this should not be unusual* is itself ... unusual.

There's much more to Collins' thoroughly entertaining article—and I already discussed the article in January (*Cites & Insights* 8.1). How do I know that? Because I tried Collins' 10-word sentence in Google…and *Cites & Insights* 8.1 comes up third in a list of seven hits. (GBS still shows zero hits.)

When I discussed this in January 2008, I didn't entirely buy the notion:

> I would note that this is probably not the case for descriptive nonfiction sentences, at least taken one at a time: After all, there are only so many ways to state a fact. (That sentence, not including "after all," appears twice in a Google search—both discussions of plagiarism—but not in Google Book Search.)

The first hit for Collins' phrase is now a post at *Althouse*, a blog by Ann Althouse, a law professor. She titles the post "Hasn't it all been said before?" and follows that with "No. Everything is actually amazingly new:" and the same stuff I quoted—properly attributed and linked, of course. The second commenter on the post argues against this proposition, saying in part:

> The argument that a similar string of words *necessarily* proves plagiarism is a statistically naive argument.
>
> There are relatively few commonly used words in English; it is highly unlikely that one can come up with a phrase or

> sentence that has not been used before, even, possibly, for the same subject matter. There are, for example, only so many ways that one can discuss Hamlet's ambivalence…

The next commenter quoted the phrase beginning "it is highly…" and responded:

> Actually it is highly **likely** that any given sentence you speak has never been used before, unless the sentence is short and about a common subject. It just seems like the same sentences get reused a lot because our brains are amazingly efficient at distilling sentences down to their core meanings, which **do** get reused regularly.

The next commenter took the direct approach, searching "Ann has too much time on her hands." No match. Another commenter searched the three phrases in the long sentence above ("There are…English," "it is highly…sentence" and "that has not…subject matter")—and didn't find matches for *any* of them.

A later commenter points out the truth behind the statistics. Even if you assume a truly tiny vocabulary, the number of combinations in a sentence gets very big very fast. If you assume a mere 1,000 words (I've seen 2,000 to 6,000 words cited as the smallest plausible vocabularies for people to communicate effectively in English), you can construct one *billion* three-word sentences, one *trillion* four-word sentences and, shall we say, an exceedingly large number of nine-word sentences. (One billion billion billion, or 10 to the $27^{th}$ power—an octillion different sentences using American wording.) As the commenter notes, "Not all sequences of valid English words are valid English sentences, but what you lose for that reason is peanuts, relatively speaking." True: If 99.9% of combinations are invalid, that would still leave 10 to the $24^{th}$ nine-word sentences…for an unrealistically small subset of English. Have a 6,000-word vocabulary? The numbers mount up a lot faster: Allowing nonsense combinations, you could have 10 million times as many nine-word sentences. I write a lot, but I won't write even a billion sentences in my lifetime (almost certainly not even ten million)—and most of my sentences are a lot more than four words long.

## Expanding the Test Cases

Somehow, my English background trumped my math background—and I found it hard to believe most non-literary sentences would be all that unique. So I thought I'd run a slightly larger experiment, using random sentences from a body of writing by an author who doesn't strive for clever phrasing—in other words, pretty ordinary sentences.

Fortunately, I could locate a writer who doesn't use many fancy words, doesn't strive for literary effect, could provide a bunch of paragraphs in machine-readable form—and wouldn't take offense at being called a writer of ordinary sentences. All I had to do was look in the mirror.

The process was simple enough. I set up a simple spreadsheet, opened Google, and started copying in the first sentence of each paragraph from an issue of *Cites & Insights* heavy on essays, where odd proper names and the like would be less likely to skew the results. I expected to find matches for at least 25% of the sentences—after all, this is commonplace nonfiction writing. (How common is that last five-word phrase? Apparently not as common as I'd expect: A Google phrase search yields zero results.)

### Skewing the research

Almost as soon as I began the process (I'd planned to search 100 sentences in all, taking up to 18 words at a time and checking Google results to see how many different authors were involved, counting up to five) I ran into trouble.

I was consistently coming up with one author: Me. Even on shorter sentences. This made no sense. I deleted a couple of sentences that used the word "liblogs." No help—and although some sentences used "blogs," surely there have been millions of sentences written using that word by now.

The ones and zeros (portions of *Cites & Insights* don't seem to be indexed by Google, although most of it is) kept on coming. After 20 or so, I started *deliberately* skewing the research toward "indistinct" sentences. I omitted sentences with proper nouns and sentences with nouns much more unusual than "librarians." I started selecting smaller portions of sentences. And I set up a parallel column, taking the first eight words of sentences and retesting those: Surely I'd get lots of matches then!

I also moved beyond that issue of *Cites & Insights* to some unedited copy for this issue (being unedited, it was even more likely to be humdrum) and unedited drafts for an early "Crawford Files" column and a "disContent" column. All the while, I was avoiding distinctive nouns and what I thought of as distinctive writing.

And still there were few matches—a few more in eight-word subsets, but even there not all that many.

### Adding more authors

After 130 sentences or so (I was fascinated enough to enlarge the sample) I decided to broaden the range of authorship a bit. That had actually happened already: A few of the sentences were quotations from blog posts.

I took a handful of well-known liblogs and tried ten sentences from each—again, avoiding sentences that were inherently distinctive because of the terminology. In doing so, I noted that Dorothea Salo's writing is distinctive even in short bursts, as are the posts of a few of the other bloggers I sampled.

I went outside the library field for one essay from *New York* (on the death of traditional publishing)—and then I tried something different: *Wikipedia*, frequently faulted for plagiarism. I took one essay that seemed like a good candidate (and included distinctive words in this case), "Jeremy Bentham," and another essay on a fairly obscure topic, "Theosophy." I did find cases that had the feel of plagiarism—but, except for the *definition* of theosophy (where an edit war seems to keep inserting a plagiarized definition), the cases of possible plagiarism seemed to be the other way around: Other websites using *Wikipedia* text without attribution. I'm not saying *Wikipedia*'s free of plagiarism; I'm saying I didn't find obvious instances in the two articles (out of several million) and 18 sentences (out of hundreds) that I tried, except for the one definition. In the end, I removed the *Wikipedia* tests from the overall sample, replacing them with 18 others from my own unedited drafts in order to maintain overall coherence.

## The Results, Round One

I tested 300 full or partial sentences—most twice. Forty-four test phrases were eight words or less; for the other 256, I also tested the first eight words.

Ten phrases out of 300—just over three percent—showed up more than once in Google, not counting attributed quotations. Here's the full list—after all, it's short!

- A funny thing happened on the way to this column
- On the other hand, it's a lot of work
- I swear I don't do this on purpose
- Let's go a little further
- Times change, and change again.
- We learn in many ways
- That time came and went
- They're free to express their opinions within reason
- That does not mean print is dead
- Improved technology cuts both ways

The first seven phrases were used by at least five different writers. The last three were used by two writers each.

Only two of these phrases are longer than eight words, and the first is an awfully convenient way to start an offbeat column (I believe it always appears as the first sentence in an article). I'll take credit for six of the ten phrases so ordinary they couldn't possibly represent plagiarism (the first, fourth through sixth, ninth and tenth).

### *The short ones*

Half of the non-unique phrases are only five words long—and, as it happens, none of the five-word phrases I tested turned out to be unique. Remember that I skewed selections toward non-uniqueness—I mean, "we learn in many ways" and "let's go a little further" (both my sparkling prose) are so ordinary they come close to cliché status.

But there were also three four-word phrases—and all three of them tested as unique:

- Delayed commentary makes sense
- This is a scattered essay
- Is blogging scholarly communication

So did all but one of the six six-word sentences:

- Formal language does not grant authority
- Personal attacks undermine reasoned arguments
- But who cares about my conclusions
- Blind posts can damage honest discussion
- I'm ignoring all sorts of context

And all eight of the seven-word sentences and initial phrases:

- Was this genuine controversy or incited controversy?
- Citations are tricky, so many different formats
- She plays us a few of the clips
- It's time common sense prevailed in Washington
- I'm a reasonably well-read, well-informed, well-educated person
- Collegiality and professionalism are perfectly fine qualities
- Blogging does have a real intellectual value
- That seems unlikely as a general situation

I wrote all six of the six-word sentences—but only one of the seven-word sentences, which came from seven different sources.

In all, 22 of the 300 test cases (7%) were sentences shorter than eight words, with another 22 (7%) exactly eight words. Two of the eight-word sentences showed up more than once, but 91% were unique. It's certainly true that most non-unique test cases were eight words or fewer (eight of ten), but also true that most short sentences were still unique (81%).

### *Length distribution*

Here's the distribution for the 256 test phrases longer than eight words:

- **Nine words**: 31 cases
- **Ten words**: 45 cases
- **Eleven words**: 30 cases
- **Twelve words**: 51 cases
- **Thirteen words**: 32 cases

- **Fourteen and fifteen words**: 25 cases each
- **Sixteen words**: eight cases
- **Seventeen words**: six cases
- **Eighteen words**: three cases.

### *A sampling of "unique" sentences and phrases*

Here are a dozen of the phrases and sentences so unusual they don't show up anywhere in Google's corpus—or if they do, it's only in the source from which I quoted and in sources properly quoting that one:

- Maybe not, but it turns out that this one works for me
- The task force has recently completed an initial draft report with recommendations
- The easiest thing to do would have been to skip the whole discussion
- He offers a few sentences that speak to what I'm getting at
- Partnerships are where you find them and what you make of them
- The other day I was walking from a meeting with a valued colleague
- It occurred to me that I'd probably be quite natural in a similar role
- People are angry and confused, searching for meaning and otherwise unclear how to respond
- There is much of interest in the specific results
- If you haven't witnessed this type of behavior in person
- I think the answer is still yes, at least some of the time
- We disagree on a number of issues—and do so agreeably

I'm reluctant to label any of these as ordinary text, since some of them come from other people. They're all clear and straightforward (which may not be ordinary at all). I'm guessing the authors would not suspect plagiarism if they happened to see these sentences in someone else's writing. (I included sentences from the following blogs in this test—and, with few exceptions, it would be easy to find the original: *The aardvark speaks, Blogwithoutalibrary, Catalogablog, Caveat lector, Free range librarian, Librarian.net, Mamamusings, LibrarianInBlack, Off the Mark, Lorcan Dempsey's blog, Open stacks, The travelin' librarian, The medium is the message, The shifted librarian, Tame the web* and *Walking paper*.)

### *Conclusions*

Common language isn't nearly as common as you might think—or, rather, "ordinary" sentences seem to be unique a great deal more often than I would have anticipated.

Is it reasonable to suggest that nine of ten sentences (nine words or longer) are unique? I have no idea. That was the case in this small sampling, deliberately excluding most sentences I thought *likely* to be unique—except that it was more than 19 of 20.

Sure, if you look at the statistics, that seems likely. But it *feels* wrong, at least to me.

Is it a matter of vocabulary? I couldn't resist that question. The 300 text samples total 3,392 words—and include 1,219 *different* words, with no normalization except for capitalization. That's a modest vocabulary.

## The Results: Round Two

Roughly five-sixths of these sentences are at least nine words long. So, I suspect, are most sentences in everyday written and spoken English. In Collins' single test, uniqueness occurred at the ninth word. What happens with the 300 samples in this run if they're truncated to eight words?

More of them show up from more than one author—35 more, for a total of 45 out of 300 tests, or 15%. (Naturally, the ten tests that weren't unique at all weren't unique at an eight-word limit.)

Of those commonly occurring briefer phrases, 28 have at least five different occurrences. I kept looking through the first hundred results (if there were that many) before giving up. Three more had four sources, four had three each, and ten had only two sources.

Some of the briefer phrases that weren't unique:

- But if I want to go back to something
- I also found it interesting that there were
- I can't tell you how exciting it is
- I come down strongly on the side of
- I don't think I would have said that
- I have to say I had no idea
- It really got me thinking about how
- Oh, and for those of you unfamiliar with
- We need to treat each other with dignity
- What is very interesting to me is that
- We have a small but excellent group of
- Fall is always a hectic time of year
- You can read the fine print yourself, but
- If you see something that is not just
- It occurred to me that I'd probably be
- So a few weeks ago we started the
- We have been having some internal discussion about

### *Some "unique" shorter phrases*

A dozen of the 255 test phrases that still showed up once (or not at all) when limited to eight words or less:

- A sort of sewing kit for my life
- Absurd and even dangerous as it may be
- All the while he has argued for a
- And, later, the clear suggestion that increased plagiarism

- As a profession how do we find, identify
- At first, I was just thrilled to see
- Copyright helps maintain a balance between the needs
- He makes the interesting point that although a
- Humans may be flawed, but we have discovered
- Mixing the old and the young, the established
- That won't reassure those who prefer to worry

*Conclusions*

The same as before, I think—although at a lower level. Even relatively short sentences seem to be unusual most of the time. On the order of 85% in this sample, and I suspect that percentage would be higher in a truly random sample.

## Uniqueness and Plagiarism

While this is only an anecdotal study, I find it mildly convincing. Our sentences are much more varied than I expected, even when we're not striving for literary excellence. (To those whom I've quoted here—always without attribution—who *do* strive for literary excellence and distinctive phrasing in each and every blog post: My apologies. It's all *good* writing, or at least I'll grant that for the 55% of these samples that other people wrote.)

What this little study does *not* show: That one duplicated sentence is evidence of plagiarism. There are, indeed, only so many ways to discuss Hamlet's ambivalence. Or are there? Searching the words "Hamlet" and "ambivalence" in Google yields a claimed 57,000 results—and I didn't spot any obvious duplications of phrasing in the first hundred.

What I believe may be true: If you're suspicious that a clumsy plagiarist has cut-and-pasted without paraphrasing, almost any medium-length sentence *may* suggest you should check further. It may be entirely innocent But it seems surprisingly uncommon for the same, say, 11-word string to show up more than once.

## Library Access to Scholarship

# OA Controversies

As we were bluntly reminded by the House Judiciary Committee, there's a gap between what we'd *like* to focus on regarding access to scholarship—and what we're *forced* to pay attention to. I'll suggest we *should* be focusing on these issues, among others (and note that some folks are doing a fine job of focusing on them):

- Success stories in institutional repositories and finding ways to make such repositories effective and sustainable.
- Investigating actual visibility and effectiveness of "green OA" through different means—that is, the comparative improvements in article impact through availability through institutional repositories as compared to subject repositories such as arXiv and PubMed, and what we can do to make the two comparably effective (if they're not now).
- Investigating and demonstrating *real* costs for gold OA and the range of such costs under different scenarios.
- A bunch of other issues, a number of them noted in "Open access issues" on the PALINET Leadership Network (pln.palinet.org/wiki/index.php/Open_access_issues)

Have I mentioned lately that the PALINET Leadership Network, my "day job," has a fine collection of articles on open access designed to help leaders get up to speed? The article noted above will guide you to the others—eight in all as of this writing.

I'd like to summarize progress in some of these areas. It's fun to write about wonderful new developments. Unfortunately, we also need to stay aware of the direct threats to open access and its advances. Here's a key truth:

- The enemies of open access have large budgets, are well organized, and have shown little reluctance to bend the truth or repeat discredited statements. And they don't give up.

Enemies? Isn't that a strong word? Well, what else would you call PRISM, to take one example? And what else would you call the forces behind September's House hearings on the NIH policy?

There are billions of dollars a year at stake—small potatoes in the world economy, but big numbers for academic libraries and scholarly publishing. If open access *really* works, it's going to shift some of those dollars. I've long held the possibly-naïve hope that some of that shift could improve the ability of academic libraries to maintain strong monographic collections and to conserve and preserve their collections and services. That's mostly why I still write about library access to scholarship.

The companies and associations now taking the lion's share of those billions don't want to give them up. The extreme cases—a small group of mostly-European for-profit STM publishers with *very* high profit margins—make noises about supporting open access, but seem intent on doing so only in ways that will assure that their revenues and profits continue to grow. They're joined by a fair number of associations that have grown to rely too heavily on journal profits to fund other association activities, forcing libraries and their institutions to subsidize the associations. Some of those associations are clearly unwilling to reduce or abandon those subsidies—and their spokespeople are in some cases willing to attack OA even more sharply than the companies making the bulk of the profits. I won't call these people useful fools; I assume they know what they're doing.

I'd accumulated some items over the past two years dealing with opposition and extremes. The second part of this essay will note and comment on some of those items. But first, there's the immediate case—attempts by publishers to get Congress to undo the NIH archiving policy that Congress mandated, weak though that policy is.

## The NIH Mandate

It's a shame the National Institutes of Health could just spring this radical mandate to seize control over research while nobody was watching. Which is to say:

- In July 2004, the House Appropriations Committee adopted a set of recommendations including one instructing NIH to develop a policy requiring free online access to articles based on NIH-funded research, no later than six months after publication of those articles in peer-reviewed journals. Note the date: July 2004, more than *four years* ago.
- In September 2004, NIH released a draft policy for a 60-day comment period.
- With full awareness of Congress, NIH released a final version of the policy to take effect in May 2005—more than three years ago. These policies allowed up to a one-year embargo and constituted a request, not a requirement. Compliance was very low—below 4% as of January 2006.
- As Peter Suber put it in early 2006, "Congress asked for a strong policy and NIH delivered a weak one." NIH also noted that 100% compliance—*all* NIH-funded papers being deposited in PubMed in a timely fashion—would cost $15 million, a tiny portion of NIH's $28 *billion* budget.
- Slowly, ever so slowly, Congress and the NIH started moving from the ineffectual request for deposit to a requirement, a mandate.
- That mandate was finally achieved, in an omnibus spending bill signed in late December 2007—still with up to a year's embargo, but with a *requirement* for PubMed deposit. Initial indications are that the mandate is working: A much higher percentage of NIH-funded papers are showing up in PubMed, sooner or later.

There were ever so many different attacks during that 3.5-year period. I suspect PRISM and the DC Coalition both have more to do with NIH than anything else. Still, after a very long period, a *very* modest mandate (that the public should eventually be able to see the results of the research it has paid for) became real.

So the publishers started using different tactics.

### *NIH's acid test (and Surprise!)*

Two posts by Dorothea Salo on *Caveat lector* on July 15 and 16, 2008 respectively (cavlec.yarinareth.net), with additional material from July 15, 2008 and July 19, 2008 posts at *Open access news*. The story involves the NIH mandate and the American Psychological Association—and a more direct publisher pushback than most of us had seen previously.

Here's what Salo excerpts from the APA policy statement as of July 15 (NOT-OD-08-033 is the NIH deposit policy):

> Authors publishing in APA or EPF journals should NOT deposit, personally and directly, Word documents of APA-accepted manuscripts or APA-published articles in PubMed Central (PMC) or any other depository. As the copyright holder, APA will make necessary deposits after formal acceptance by the journal editor and APA….
>
> In compliance with NOT-OD-08-033, APA will deposit the final peer-reviewed manuscript of NIH-funded research to PMC upon acceptance for publication. The deposit fee of $2,500 per manuscript for 2008 will be billed to the author's university per NIH policy. Deposit fees are an authorized grant expense. The article will also be available via PsycARTICLES.

Salo's summary:

> If you want to comply with the law demanding deposit in PubMed Central for an article we're publishing, you pay us $2500. Do not complain, do not pass go, do not deposit the article yourself (which is free).

She calls this the acid test for NIH, and says that unless NIH acts, other publishers will set similar fees. "There's no risk in it for them until you create one." Peter Suber came down hard on the APA decision, saying, among other things:

> No author or author-sponsor should ever have to pay a fee to deposit an article in an OA repository… [The NIH] certainly doesn't *require* publishers to charge fees. The APA is simply being dishonest when it says that it will bill its fee to universities "per NIH policy"… The foulness of this policy wouldn't matter if NIH-funded authors simply steered clear of APA journals…

And here's where it gets interesting. *One day later*, the APA statement was gone. In its place:

> A new document deposit policy of the American Psychological Association (APA) requiring a publication fee to deposit manuscripts in PubMed Central based on research funded by the National Institutes of Health (NIH) is currently being re-examined and will not be implemented at this time. This policy had recently been announced on

APA's Web site. APA will soon be releasing more detailed information about the complex issues involved in the implementation of the new NIH Public Access Policy.

APA will continue to deposit NIH-funded manuscripts on behalf of authors in compliance with the NIH Public Access Policy.

Salo's comment:

I hear through the grapevine that librarians were going to faculty who edit APA journals and asking whether they liked what they saw. If that worked, which is admittedly still to be determined, it suggests that such outreach should be standard procedure in cases like this. Find the editors on your campus, lay out what's going on, ask whether it's all right by them.

A few days later, Suber discussed the "new interim policy"—which is spelled out in a little more detail on the current site. To wit, APA will deposit the final peer-reviewed Word document (not the published article) for NIH-funded research "at the appropriate time," presumably after a year—but will deposit the *published* article *immediately* for research funded by Wellcome Trust…but then, Wellcome is paying $4,000 per article for that immediate access.

It's unfortunate (but predictable) that Stevan Harnad saw this as another excuse to attack the NIH mandate because it uses PubMed rather than institutional repositories—as always, Harnad sees One True Way and does his best to strike down any heresy, no matter the damage to open access as a whole. (Harnad appears to *defend* APA's action—and accepts APA's assurance that it didn't say what it *explicitly* said in the July 15 statement, withdrawing its former "green OA" policy for NIH-funded articles.)

This was a sideshow—an instructive sideshow and one that *might* show the power of concerted action by scholars, but a sideshow nonetheless. This year's main event came in September 2008…

All the notes that follow come from a series of posts at Peter Suber's *Open access news* for September 5, 2008 through September 16, 2008 (www.earlham.edu/~peters/fos/), although some of those posts cite other sources. The full series of posts and other linked items tell the story in far more detail than I'll offer here.

### *Publishers go to Congress to undo the NIH policy*

Suber quotes extensively from a September 5 story by Andrew Albanese at *Library Journal*. Excerpts:

In less than a week…the Subcommittee on Courts, the Internet, and Intellectual Property of the House of Representatives' Judiciary Committee is scheduled to hold a hearing on what sources tell *LJ* is a legislative attempt to redress publishers' concerns that public access policies —namely the recently enacted policy at the National Institutes of Health (NIH)— conflict with copyright and intellectual property laws…

The legislative hearing comes after publishers succeeded in adding a key phrase to the NIH public access mandate just before the bill's passage in December, 2007—that the NIH policy be implemented "in a manner consistent with copyright law." As *LJ* reported then, that simple phrase appeared to position publishers for a possible legal or legislative challenge to the policy.

In recent months, the possibility of a legal or legislative challenge began to seem almost certain…

Anticipating such a challenge, officials at SPARC and the Association of Research Libraries, however, have strongly denied that the NIH public access policy conflicts with copyright, last year preparing a memo of their own…

Suber links to a current SPARC analysis of the NIH policy and copyright law (www.arl.org/sparc/bm~doc/ nihpolicy_copyright_july2008.pdf). He notes careful wording in one of the anti-NIH complaints—not that the policy actually infringes copyright (it doesn't) but that it infringes a right publishers *don't have* for NIH-funded research. Here's part of Suber's rewording of publisher complaints:

"OK, the policy doesn't violate the letter of copyright law, but it violates the spirit, which is that our ability to profit from research we didn't conduct, write up, or fund should not be put at risk just so that publicly-funded research can be made more useful, by reaching everyone who can make use of it, or just so that taxpayers don't have to pay twice for access… [T]he spirit of copyright law is that [authors] should transfer all of their rights to publishers. We've grown to depend on it… The government should protect us from risks created by new new and better ways of doing things. It violates the spirit of copyright law for a government agency like the NIH to put the taxpayers' interests ahead of our private interests as an industry."

### *The bill and the hearing*

The Fair Copyright in Research Works Act (HR 6845) was introduced on September 9 and the subcommittee held a hearing on September 11. Briefly, the bill would prevent any mandate similar to the NIH policy for any "extrinsic work"—which appears to mean not only any research funded *in any part* by anybody other than the Federal government, but also any research where there's "meaningful added value" from anybody other than Federal agencies. It's not just NIH; it's *any* Federal agency—and it pretty much eliminates those agencies' rights to require access licenses.

Suber's initial take is that the bill didn't seem likely to be passed this year, given elections and all—but it's a "stalling tactic" that diverts OA energy and could "change the narrative" so that media and policymakers talk about "changes in copyright" rather than increases in access.

One blogger (Karen Rustad) may have gotten it in one as to why this bill even earned a hearing, given that the NIH mandate involves contract law and doesn't *at all* affect copyright:

> The only reason this has even made it to public comment (I think) is a bunch of representatives feeling slighted because a bill passed Congress without going through their committee (the Subcommittee on Courts, the Internet, and IP–the measure was part of an Appropriations bill, so it went through that committee). The grumbling at the opening of the session about how important their committee is, prestige of the Appropriations committee be damned, rah rah rah, I think bears this out. So the representatives have been receptive to the patently ridiculous argument that the NIH mandate *changed* copyright law and, thus, should have fallen under their purview.

The subcommittee involved includes "Hollywood Howard" Berman, who I've mentioned in various copyright-related contexts, and Rep. Conyers (who introduced the bill and is also fairly consistently in favor of increasing the reach of copyright under all circumstances). No party issues here: extreme copyright advocates are on both sides of that line.

Later that day, another *OAN* post notes statements from our friends the DC Principles Coalition and AAP's PSP division and from the Copyright Alliance. An excerpt from the DCPC/PSP letter:

> A recent congressional mandate at the National Institutes of Health (NIH) forces publishers to surrender their copyrighted scientific journal articles for free public access twelve months after publication and sets a dangerous precedent. This mandate in effect reduces copyright protection for this important class of works to only one year…

It does no such thing, to be sure, but that's never bothered these groups. The letter from Patrick Ross of the Copyright Alliance is worse:

> The mere fact that a scientist accepts as part of her funding a federal grant should not enable the federal government to commandeer the resulting research paper and treat it as a public domain work....[T]aking the scientist's copyrighted interpretation of the data is not fair to other funders, and it is certainly not fair to the publisher. A publisher improves the work through a rigorous peer review process and develops it for publication. Authors and publishers don't need the feds playing Rumpelstiltskin by returning after a year to take their children away.
>
> That publisher has earned the right as a copyright owner to pursue a return on his investment, a pursuit made more difficult when its copyright term is essentially reduced to one year.

As Suber notes, "surrender" is a dishonest description, and indeed the NIH policy has no effect on the published version. Suber traces "surrender" back to PRISM and its "dishonest advocacy." But Suber also notes "commandeer" as an escalation of rhetorical excess (noting "It would be at least as accurate to say that the traditional publishing model commandeers publicly-funded research, and holds it for ransom") and the deceptive use of "public domain."

Still later on September 11, Suber posted excerpts from stories on the hearing in *Library Journal Academic Newswire*, quoting some interesting notes from statements. Allan Adler (AAP) says "government does not fund peer-reviewed journal articles—publishers do." That's wonderful sophistry, since what publishers fund is the copy-editing and markup (most peer review is done for free), not either the research that led to the articles or the writing of the articles themselves. When a SPARC representative noted that peer review is done for free, with the main cost to publishers being "sending some emails," Martin Frank of the American Physiological Society made a rather astonishing statement. He said APS spent $13 million a year to publish 14 journals (I wonder whether that's expenses or revenues, but never mind) and that "sending those emails" accounts for about 20% of the publishing costs—or $2.6 million a year. Man, that's a lot of email, or a lot of *something*. In his statement, he also said "The NIH has become a publisher."

The copyright office came in on the side of the publishers. Ralph Oman said that in his opinion the NIH mandate would "destroy the market" for commercial scientific journals (which should be entirely out of his domain) and cause a "dilution" of copyright. He even referred to "the hairy snout" of government, saying it should be kept out of science publishing—but, oddly, nobody seems to think that NIH should keep its "hairy snout" out of research itself, which costs many times as much as the publishing (NIH figures $300,000 per article). A SPARC representative and the head of NIH both defended the mandate—and the SPARC executive director offered a personal anecdote, noting that her five-year-old son was diagnosed with diabetes and she was able to access worthwhile information thanks to NIH's policy.

### *Stuff from September 12*

There was a *lot* of activity the day after the hearing—*OAN* has seven posts throughout the day, some of them citing multiple sources. A few highlights, if that's the right word:

- A *Chronicle of Higher Education* report uses dramatic terms: "A life-and-death battle is going on over public access to federally financed research—life for taxpayers and many scientists, and death for publishers. Or so each side claims...." The writer notes who's missing from the hearing: Scientists—but 33 Nobel Prize winners submitted an open letter calling the publishers' move "wrong" and NIH's policy "enlightened." Martin Frank is quoted saying "Articles should not be taken from those of us responsible for their creation"—which is an interesting way to put it, since naïve citizens might assume that researchers create the articles, not journals. (Suber calls Frank's claim "breathtakingly one-sided.")

- The American Chemical Society explicitly supported what was now being called the "Conyers bill." The writeup on this story makes it clear that *all* peer-reviewed articles would be "extrinsic work" even if the research is wholly funded by Federal agencies—because the peer review and publication process represents non-government funding. Neat.
- *Science Magazine* covered the hearing, noting that NIH says compliance has risen to 56%. The report includes a statement from Jonathan Band (representing ALA) noting the bill's sweeping provisions as a fatal flaw—and this startling pair of sentences: "Representative John Conyers (D–MI)...questioned the need for the policy when the public can already obtain the papers through a subscription or at a library. Moreover, most journals make their content free after 12 months." If the second sentence is true, there is no plausible reason for journals to object to the NIH mandate; the first is the kind of "let them eat cake" argument you wouldn't expect from a Michigan Democrat, and ignores the fact that "a library" really only means some larger university libraries in all too many cases.
- The American Association of University Presses, AAUP, sent a letter favoring the Conyers bill. The letter's heavy on the phony "diminishing copyright" claim and, frankly, brings nothing new to the table. Later, the executive director of Rockefeller University Press wrote to disagree with AAUP's stance and say the press "strongly opposes your efforts to overturn the NIH mandate."
- A piece in *Government Executive* clarifies the turf warfare at play—Conyers' outrage that the House Appropriations Committee didn't consult with his subcommittee before "pushing through" the mandate (after a mere three years of discussion and feedback). That piece says Howard Berman did not publicly endorse the bill—and a sidebar suggests that he might actually oppose it. (Suber notes that, since the NIH mandate has *nothing* to do with copyright, there was no reason for the Appropriations Committee to consult Conyers' group. You think?)
- Apparently, Rep. Bob Goodlatte (R-VA) has an interesting idea: NIH should *only* post *unreviewed* articles to PubMed Central with a disclaimer. I didn't say it was a *good* idea—and this was a case where Martin Frank and NIH's Zerhouni actually agreed the move would be unfortunate ("disastrous" was Frank's term), given the high rate of rejections in scientific journals.
- A post by Michael Carroll notes that the Conyers bill would do more than reverse the NIH mandate: It would implicitly amend existing law and repeal longstanding contractual provisions for Federal contracts. He notes that a group of 47 law professors, all copyright specialists, sent the committee a letter noting that there's no basis for the claim that NIH's mandate is inconsistent with copyright. The letter notes that the NIH mandate is part of its funding contracts, not a copyright issue—and that the author chooses to accept NIH funding (and the contract) long before a publisher can enter the picture, so it's not possible for the mandate to take intellectual property away from the publisher.

### *The close—for now*

By September 16, it was fairly clear that the Conyers bill wasn't going anywhere *for now*. A *Library Journal Academic Newswire* report quotes Howard Berman saying the bill would be held "until at least next year."

Over the following days, some of the letters to the House panel and other reactions popped up. A group of mostly library folks (AALL, ALA, ACRL, ARL, Public Knowledge, SLA, SPARC and others) noted the worth of the NIH policy and clarified that it did not affect copyright law (enclosing the SPARC policy brief noted earlier).

*Ars Technica* had its usual lively (and frequently dead-on) reporting, including Howard Berman's apparent shock when Elias Zerhouni informed him that NIH "hands out $100 million a year to grant recipients specifically to cover the cost of publishing their results" (remember: a higher percentage of toll-access journals charge author-side fees than do open access journals), supposed "surprise…that authors were not paid by publishers for the transfer of copyright," and this:

> If anyone was thinking that policies related to publicly funded scientific research were free of politicking and rampant self-interest so frequently involved in the copyright and intellectual property battles, the hearings would have erased them....

Paul Courant compared the Conyers bill to the Clear Skies Act, "an odious piece of corporate welfare wrapped up in a friendly layer of doublespeak," noting:

It would make it illegal for U.S. government agencies to seek any rights at all in the research that they fund. This is anything but fair. Indeed, it is manifestly unfair to the taxpayers who ultimately pay for the research, and on whose behalf the research is conducted…

The people of the United States pay good money to learn about the world. It would be a travesty if Congress decided that the interests of a few publishers were more important than the research investments of the American public, and that's exactly what this bill would do.

There was lots more—almost all of it decrying the Conyers bill. But it's never really over. The Alliance for Taxpayer Action called on citizens to contact their Representatives and Senators to make sure HR6845—and the provisions in the bill—are defeated. The call (quoted at *OAN* on September 18, 2008) spells out what's wrong with the Conyers bill. A comment notes that the language of the bill could still be inserted into other legislation.

## Other Opposition

Let's look at some other opposition to open access (or commentaries on opposition) over the last year or so.

### *Death knell for peer review?*

Richard Poynder published "Open access: death knell for peer review?" on October 15, 2006 on his blog, *Open and shut?* (poynder.blogspot.com) It is, as with most of Poynder's solid journalism, a carefully done, thoughtful and fairly long piece. (Well, not *C&I* long, but long for a blog post.)

He notes the argument made by publishers that OA threatens peer review. He considers the two forms of OA: "green" OA, where the published paper is in a traditional journal with traditional peer review—and "gold" OA, where the paper is in an OA journal…with traditional peer review.

Since both methods still require that papers are peer reviewed, OA advocates point out, publisher claims that making research OA necessitates foregoing the peer review process is factually inaccurate.

I'm not really an OA advocate—at least not a strong one—and I have *never* been able to find any truth in assertions that OA would weaken peer review.

There is, however, a second strand to publishers' claims that OA threatens peer review. If OA is forced on them, they say, they will not be able to survive financially, either because they will discover that there is no stable long-term business model for OA publishing, or because the increasing number of papers researchers post in institutional repositories will cause academic institutions to cancel their journal subscriptions. This poses a threat to peer review, they add, since if publishers exit the market there will be no one left to manage the process.

However, these claims are also rejected by OA advocates, who argue that most publishers have already accommodated themselves to self-archiving. Indeed, they add, there is no indication at all that self-archiving negatively impacts journal subscriptions. Nor is there any reason, they say, to believe that a sustainable OA business model cannot be found.

In a way, this could settle it: There is simply no objective basis for the claim that OA would weaken peer review. But Poynder takes it one step further:

But supposing publishers are right, and OA does eventually cause peer review to be abandoned? Would it matter?

He notes objections to the conservatism of peer review and even publisher statements calling the process flawed.

In fact, it seems that the most that can be said of peer review is that we have failed to come up with anything better. Following the decision by *Science* to retract papers it had published by Dr Hwang Woo-suk — after it was discovered that he had faked claims that he had obtained stem cells from cloned human embryos — for instance, publications consultant Liz Wager said of peer review "it's a lousy system but it's the best one we have."

Poynder then notes moves by some publishers to "open review," where the names of the reviewers would be known to authors, and the "more radical approach" of *PLoS ONE*, where papers undergo a less rigorous peer review and are subject to post-publication review on the web.

*PLoS ONE* referees are asked to answer a simpler question than that asked by traditional peer review. That question is: "Has the science in this paper been done well enough to warrant it being entered into the scientific literature as a whole?"

But you can be even more radical—as in *Philica*:

*Philica* has no editors, and papers are published immediately on submission—without even a cursory review process. Instead, the entire evaluation process takes place after publication, with reviews displayed at the end of each paper.

As such, the aim of the review process is not to decide whether or not to publish a paper, but to provide potential readers with guidance on its importance and quality, and so enabling particularly popular or unpopular works to be easily identified.

Importantly, argues *Philica*, its approach means that reviewers cannot suppress ideas if they disagree with them.

The evaluation process involved "recursively weighted reviews." Unlike *PLoS ONE*, there are no author-side fees, since overhead is nominal. (*Philica* uses a filter: Only academics are allowed to submit reviews.)

There's a *lot* more to Poynder's 4,200-word post. Worth reading.

### *Hearing from IASTM*

The International Association of Scientific, Technical & Medical Publishers (*www.stm-assoc.org*) represents the biggest STM journal publishers and some others. Last October, it published a position paper on why publishers seek copyright transfer: "to ensure proper administration & enforcement of author rights."

That's right: It's not for the publisher, it's *for the author*. Bullets point out that authors "are rarely in a position to defend themselves against infringers, plagiarists, pirates and free-riders" and that everyone benefits from the broadest possible dissemination (which, IASTM claims, is facilitated by "the publisher"). There are other possible benefits, some probably legitimate (although some, such as subsidiary rights management, simply *do not require* turning over the copyright—I say as the author of several books and many articles where I did *not* turn over copyright but *did* authorize the publisher to handle subsidiary rights).

I suspect most scholars who write journal articles aren't too worried about copyright infringements, "pirates" or "free riders"—they're not getting paid for the articles anyway. They want dissemination, impact and credit: Plagiarism is indeed a concern. Has IASTM or any publisher ever provided proof that they've acted to prevent plagiarism? If so, this claim might be more meaningful. As for broadest possible dissemination—well, it simply beggars belief to claim that a closed arrangement (turning over copyright) results in broader dissemination than, say, a typical freelance writer's contract (*for which the writer is paid*) or true OA. I can post all of my *paid* magazine columns, for free, a year after they're published—because I hold copyright and assigned typical freelance rights to the publisher.

In April 2008, the organization published *An Overview of Scientific, Technical and Medical Publishing and the Value it adds to Research Outputs*, a 16-page treatise. You can download it from IASTM's website.

It's an interesting piece of work. I'm surprised to see publishers suddenly claiming they provide preservation services; that's not something STM publishers have traditionally seen as their purview. Nonetheless, the benefits section seems good. Then we get to "There is no publishing 'for free'" and some tricky commentary. Somehow, the full costs of "editorial office management systems" become assigned to peer review management and we're told that editorial work within universities is typically charged back to publishers.

Then there's "STM Publishers and the Goal of Open Access." Right off the bat, OA is called a "visionary goal" and IASTM asserts a rigid set of required characteristics for any new business models—some of which essentially says "we have to make our money, no matter what." They throw out fairly high numbers for the costs of IRs—and when there's a firm statement, as from NIH, they undermine it with something like "it is widely believed that their estimates of current and projected costs may be a considerable underestimate and do not include important elements such as staff time." The source? "It is widely believed." They headline the ARL Spec Kit that found one startup cost estimate as high as $1.8 million—the extreme end of a range gets the featured number.

The piece goes on to argue against systematic or interlinked self-archiving—in other words, green OA is fine as long as it's wholly ineffectual. We're also informed that embargo limits are unacceptable because they don't fully secure subscription revenues (while, in a nicely ironic touch, also not fully realizing the OA goal of immediate availability).

Maybe I'm misinterpreting. Maybe this is an objective overview of the topic—not a nicely done case of special pleading against *any* effective form of OA and for assured profits. I just find it hard to read that way.

### *The science and the say-so*

I've generally avoided commenting on Joseph Esposito's writing. I slipped twice—once in September 2004, discussing a *First Monday* article, and again in Spring 2006 only because his article was part of the return of *Journal of Electronic Publishing*. His articles consistently make me so angry, because of the writing and ideas, that I'm never sure I can comment on them in a reasonable fashion—so I've learned to not read them, even if they're in journals I otherwise respect (like *JEP*, for example).

Alma Swan doesn't have that privilege, particularly since Esposito mentions her work in some of his writing. She wrote this post on December 21, 2007 in *OptimalScholarship* (optimalscholarship.blogspot.com), responding in part to a post on Esposito's blog.

In one case, Esposito states that OA proponents imply that librarians are stupid—because some proponents say libraries won't necessarily cancel subscriptions to journals whose contents are available via OA.

> But that misses the point: we say that cancellations won't necessarily occur because that is what we observe, in real life. It is true that it is somewhat perplexing and seems to fly in the face of logic. Why would you, in these days of straitened circumstances for libraries, continue to pay for a journal whose articles are available for nothing?

I'll admit I also find this perplexing—and don't believe it to be likely for the long run. Still, the example is clear enough: arXiv has contained the contents of many journals in some fields of physics for 15 years, and the journals haven't seen mass cancellations. (Swan points out that arXiv isn't just preprints: More than half of the articles are postprints.)

> There are a number of straightforward reasons…for preferring to continue to subscribe to journals–journals are more than just articles and contain other types of content that people want to read; they contain the final polished-up versions of articles whereas OA versions are simply the author's final product; there is no guarantee that every article from a journal will be made OA by its author. Some of these may not hold up forever. We will start to see which journals have true added value–that is, something that customers will pay for–and which are just a collection of articles: the marketplace will reveal that. There are also other reasons, ones not so straightforward and certainly not so easy to describe. They are to do with allegiances to certain publishers, particularly specific society publishers who are viewed as 'the good guys' and thus worthy of loyalty; they are to do, partly, with the sorts of deals that publishers are prepared to offer in every individual case; and then they are very much to do with the views of faculty, without which no librarian makes a final decision on what to cut and what to reprieve. And faculty have very strong views on these things, not all of them based on logic or evidence. Even high-energy physicists have feelings.

Then there's a statistic, one Esposito belittles. It came from a 2003 study commissioned by JISC and included a "What would you do…" question of the sort Swan prefers not to ask. The question was what respondents would do if their employer or funder required them to make their work open access, and there were three options: Comply willingly, comply reluctantly or not comply.

Here's Esposito's comment:

> It was found that 81% of researchers say that they would comply with mandates. Now, what does this prove exactly? More than 81% of Americans comply for the most part with the U.S. Tax Code, but that is hardly indicative of support for the current administration or the way tax monies are spent. What it does reveal is a healthy respect for the punitive powers of The Man. In OA circles, however, a forecast compliance with a mandate is viewed as the equivalent of democratic support.

But, as Swan points out, one rule when using the work of others is to be accurate in citing that work. In fact, the results were that 81% said they would comply *willingly*. Another 14% said they would comply reluctantly. (Five percent said they would not comply.) So Esposito's comment only makes sense if he uses 95%, not 81%.

Swan agrees that the survey only provides a starting point. Where's the testing? Turns out that's been done too, by Arthur Sale, measuring the amount of material being deposited in Australian university repositories under different conditions. The results were as predicted.

There's another long section having to do with the "open access advantage"—the likelihood that OA will enhance visibility, access and citations—but you'll have to go read Swan's post (and Esposito's) to see what's going on there.

### *The American Chemical Society and Open Access*

That's the title of a Viewpoint by Bob Michaelson (Northwestern) in the Winter 2008 *Issues in Science and Technology Librarianship*, one of ALA's first gold OA publications (www.istl.org). It's short and clear, and well worth reading. Michaelson isn't a hot-blooded OA supporter; he even says "it is not yet clear how Open Access (OA) publishing may find a business model." But he recognizes its importance and says, "We must therefore sustain a serious conversation among all players--researchers, funding agencies, libraries, and publishers--about potential implications, both beneficial and detrimental, of various sorts of OA."

> Unfortunately, serious conversation is ill-served by some publishers' strategies, including, regrettably, those pursued by the American Chemical Society.
>
> Editorials in *Chemical & Engineering News* as far back as 2004 denounced OA as "socialized science"--whatever that is supposed to mean. In 2005 Nobel Laureate Richard J. Roberts published an open letter announcing his resignation from ACS out of disgust at the Society's opposition to OA.

Michaelson's particular scorn is reserved for the Dezenhall episode—AAP/PSP's hiring of the aggressive PR man, his advice that they focus on slogans (regardless of truth), the founding of PRISM and the disavowal of PRISM by several leading university presses.

> If the ACS regrets its association with PRISM's misstatements, they don't show it. After a long legislative fight, the National Institutes of Health (NIH) was able to secure passage of a mandate on open access for NIH-funded research within a year of publication… But as reported in *LJ Academic Newswire*, the ACS threatens a legal battle. Executive Director Madeleine Jacobs makes the claim, according to *Chemistry World*, that "the policy would result in conflicts with copyright law and intellectual property rights," resurrecting the claim that it could "interfere with scientific peer review" and adding that it would "adversely affect the sustainability of scientific journals." It seems to me that all of these claims are nonsense. *LJ Academic Newswire* notes that the library community refuted ACS's copyright claim in July 2007. As Peter Suber has repeatedly discussed, peer review is entirely compatible with open access.
>
> If the ACS is to be a party to discussions of OA, they must stop getting their policy advice from PR flacks and start making rational contributions to the discourse.

Otherwise they will continue to poison the waters, and deservedly will be accorded no credence.

An excellent commentary. Will ACS pay attention? That seems unlikely, based on a long track record.

## Difficulties and Extremes

I'll close with a few other items on difficulties and extremes, with the hope that the next episode of LIBRARY ACCESS TO SCHOLARSHIP can focus more on the positives—such as considerable growth in OA publishing.

### *Open access fundamentalism*

Another one from Alma Swan's *OptimalScholarship*, this time posted May 7, 2007. (She doesn't post that often—so far, 11 posts in 2007 and three in 2008—but when she does, she has something worthwhile to say.) This post was the very first post on the blog, and is far more than an "I'm here!" announcement.

Basically, Swan says she's arrived because an Outsell person has called her "a fundamentalist" and "shrill." "And what they say is that if people who disagree with you start calling you names, then they are taking you seriously." (I'll have to remember that.)

> Now first, let me say that I really like the 'fundamentalist' bit. I have always strived to avoid the superficial, speculative and emotional end of the spectrum and to base the views that I hold on facts, so far as I am able. So yes, I do fundamentalism. The 'shrill' label sits less easily, though.

That's an interesting reading of "fundamentalist," and I rather like it. She notes that the Outsell person, David Worlock, was critiquing an invited essay Swan wrote for *American Scientist* on how OA can advance science. What made her essay shrill?

> I was asked in my essay to define the ways in which Open Access advances science and there I was, thinking that I'd done so in moderated terms, supporting each point I was making with good data and reasoned argument. I detailed four ways in which science is advanced by Open Access: it enables greater visibility and, as a result, impact; it moves science along more quickly; it enables new 'Web 2.0' semantic technologies to work on scientific output, generating new knowledge by data-mining and text-mining scientific output in the vast single information space that Open Access provides ; and it enables new tools that can measure impact and effectiveness in brand new ways, a boon to research managers and funders across the world.

Apparently Worlock claims publishers will do OA *better*: "As publishers move to contain, embrace and even capitalise on the access and availability issues, they are doing so in ways that save time and energy for researchers whose concentration is upon the science involved, and its communication to small, close-knit communities of fellow workers whom they reach at conferences and via e-mail links." Of course, OA is about far more than those close-knit communities; emailed preprint copies can pretty much handle them. OA is about *open* access, not just access for the inner circle.

Swan portrays Worlock as suggesting that gold OA is nearly stagnant, which is certainly not true. She also notes a cute finish to Worlock's article, where he says you can obtain Swan's article for $12 if you're not a member of the society. Which is true—but omits the fact that the article is also available in HTML form for free (and always has been), that it's available on her institution's IR, that it's on her consultancy's website…all for free. "Must have just slipped his memory."

So, of course, I had to go read Swan's article to test for signs of shrillness. You could read it also: www.keyperspectives.co.uk/openaccessarchive/Journalpublications/American_Scientist_article.pdf.

Whether I read charitably, closely, or with intent to do damage—I will be damned if I can find anything shrill, unreasonable or even particularly opinionated in the article. It's rife with proof for claims and has a remarkable amount of data for such a short piece. But **shrill**? I'm sorry, but I just don't see it.

### *"open access" is not good enough*

That's how Peter Murray-Rust titled a June 10, 2007 post at *petermr's blog* (wwmm.ch.cam.ac.uk). Excerpts:

> I have ranted at regular intervals about the use of "Open Access" or often "open access" as a term implying more than it delivers. My current concern is that although there are are tens of thousands of theses described as "open access" I have only discovered 3 (and possibly another 15 today) that actually comply with the BOAI definition of Open Access. The key point is is that unless a thesis (or any publication) explicitly carries a license (or possibly a site meta-license) actually stating that it is BOAI compliant, then I cannot re-use it. I shall use "OpenAccess" to denote BOAI-compliant in this post and "open access" to mean some undefined access which may only allow humans to read but not re-use the information I do not wish to disparage the important efforts to making scholarly information more widely available, and I applaud the general direction and achievement of the groups below. I appreciate that the copyright of historical content normally is held by the student author and it's certainly very valuable to have "access" to it. But it is not OpenAccess. And unless specific policies are put in place to add specific BOAI-compliant licenses then future theses will also be non-compliant…
>
> By contrast let's look at "Open Source" which applies to software and has been highly successful in liberating the field. It's very widely used in academia and elsewhere…
>
> In general the term "Open Source" is completely self-explanatory within a large community. I can describe my

> software as OS and everyone understands what I mean. There are some licenses (e.g. GPL) which require additional freedoms but they don't invalidate the above. By contrast if someone describes something as "open access" it simply means that I may--as a human--and at some arbitrary time in human history--read the document. It does not guarantee that I can save my own copy, that it will be available next week, that it will be unaltered in the future or that versions will be tracked, that I can create derivative works, that I can use machines to text- or data-mine it
>
> So I believe that "open access" should be recast as "toll-free"--i.e. you do not have to pay for it but there are no other guarantees. We should restrict the use of "Open Access" to documents which explicitly carry licenses compliant with BOAI.

Peter Suber responded in some detail, and the recent suggestion of "gratis" and "libre" may relate to this confusion. The problem here is attempting to narrow OA to include only completely reusable material. For the purposes most often touted by OA supporters—universal access to research papers, greater impact, etc.—open *readability* is the key. Data mining is a much different issue—and, interestingly, Stevan Harnad (who commented basically saying that Murray-Rust should go ahead and do some of the things he believes to be ruled out) *rejects* derivative works as part of OA.

Murray-Rust is "arguing that the level of emphasis throughout the community should be higher." That's fine, but it raises the question of whether the best, in this case, is the enemy of the good. Insisting on the right to datamine and to create derivative works, and the insistence others have added that commercial reuse shouldn't be limited, leaves only attribution between OA and the public domain—and with datamining, I have to wonder about attribution.

The proposed distinction between gratis OA and libre OA (discussed in *C&I* 8:8, August 2008) may not quite achieve Murray-Rust's aims—but it avoids the considerable harm to OA as a movement that could come from saying you can't call it Open Access unless *all* permission barriers are removed.

### *Open, online journals ≠ PDF ?*

To some extent, this is a related item—Andy Powell's August 6, 2007 post at ***eFoundations*** (efoundations.typepad.com/efoundations/). Since Powell's blog has an explicit CC "BY" license and it's a concise post, I'm quoting the entire post:

> I note that Volume 2, Number 1 of the *International Journal of Digital Curation* (*IJDC*) has been announced with a healthy looking list of peer-reviewed articles. Good stuff.
>
> I mention this partly because I helped set up the technical infrastructure for the journal using the Open Journal System, an open source journal management and publishing system developed by the Public Knowledge Project, while I was still at UKOLN--so I have a certain fondness for it.
>
> Odd though, for a journal that is only ever (as far as I know) intended to be published online, to offer the articles using PDF rather than HTML. Doing so prevents any use of lightweight 'semantic' markup within the articles, such as microformats, and tends to make re-use of the content less easy.
>
> In short, choosing to use PDF rather than HTML tends to make the content less open than it otherwise could be. That feels wrong to me, especially for an open access journal! One could just about justify this approach for a journal destined to be published both on paper and online (though even in that case I think it would be wrong) but surely not for an online-only 'open' publication?

I use PDF for *Cites & Insights* to preserve typographic and layout integrity. I suspect journal publishers may do the same—and that, as with *Cites & Insights*, their layout and work flow may yield perfect PDFs readily, where workable HTML/XML might take quite a bit more effort. Indeed, the first comment (from Dorothea Salo) suggests that point: "Go find 'em a workflow that produces good HTML as well as PDF, and I'm sure they'll sign right on."

BioMed Central notes that they use both PDF and HTML—and that more than 90% of accesses to full-text articles are for the PDF versions. Cornelius Puschmann offers a long comment coming to PDF's defense for several good reasons: e.g., HTML's interpreted nature "gives you a virtual guarantee that it will render incorrectly in somebody's browser," printing HTML is still horribly inconsistent, PDF is arguably more portable for complex documents, insisting that online journals should shun PDFs "means giving readers less reliability when it comes to printing and file mobility"—and there are lighter-weight PDF readers than Adobe. (My feeling is that Reader 8 is much less demanding than earlier versions.) Others come to PDF's defense as well—including an OA journal editor who notes the extra time required to convert manuscripts to HTML. That editor notes that most readers appear to want PDFs.

I must admit that it pains me to see the number of times the rare full-issue essays in *Cites & Insights* are downloaded in HTML rather than PDF, because I suspect they're getting printed—and they use at least 50% more print space in HTML form. I'm still of two minds about providing HTML versions of essays, particularly since I can't afford the time to make them as good-looking and readable as the PDF versions. Word2007's filtered HTML is better than Word2000's, and that was good enough for my purposes. I admire Tom Wilson's ejournal—but I reject the idea that *C&I* is stone age or that my issues damage the Internet. Of course, I don't claim full (libre) Open Access, and it's somewhat irrelevant because these aren't scholarly research articles.

## Retrospective

# Pointing with Pride, Part 7

I'm trying to make these shorter—just offering some tidbits and highlights along the way.

## June 2001: Number 7

Let's start with a little silly-season stuff, namely this tidbit from TRENDS & QUICK TAKES under the headline "Get yourself an outfit and be a cowboy too!":

> Microsoft wants most of us to think we're project managers. Big two-page ads in computer magazines have the following large type: "Have you answered more than 10 questions today? If so, then you're a project manager." Thousands of technical support people, help desk staff, and reference librarians will be mighty surprised by this assertion.
>
> It's simply, really. Microsoft wants you to buy Microsoft Project 2000. Why buy a project scheduler if you're not a project manager? It's a little like the new definition of "ebooks" that includes print-on-demand books (and in a few years will probably include any book for which the text was originally in digital form): the easiest way to expand a market is to redefine it.

Think that was fun? Here's Dick Brass, Microsoft's "chief evangelist for electronic books," in an article at *Industry Standard:*

> When he spoke to senior executives at a big wood and pulp company he said, "I see dead people everywhere, and they don't know they're dead." The Microsoft timeline has a "history of printing" that ends with the final paper edition of the *New York Times* being printed in 2018—you may have seen that date in some short-lived MS ads.
>
> "Twenty years from now, 90 percent of everything published will be published electronically." That's Brass' claim. He goes further, however, in a curiously paradoxical statement. "Literacy will spread. Poverty will retreat. There'll be no village in India or Africa too poor to have a library equivalent to the greatest universities in the world." And yet, Microsoft intends to make sure that its ebooks are fully protected, with every use paid for. How will these incredible libraries come about?
>
> Brass goes on to say, "It's like 1908 in the automobile industry. Twenty years later, it was hard to find a horse in a major American city. The same will be true for books." Read that carefully: *by 2019, it will be hard to find a [print] book in a major American city*. Short of massive thermonuclear war, this vision sounds absurd. But that's what the man said, and he's a Microsoft hotshot.

The curious thing is that the "90 percent" figure might be right, depending on your definition of "published"—and that doesn't reduce the role of print books, or print in general, at all. Anyone want to bet on finding it hard to find a [print] book in a major American city by 2019? Most of the most avid ebook supporters wouldn't take that bet; they understand complementarity.

## February 2002: Number 17

I tried to offer "monophone comments on a trilingual econference," a series of articles and virtual discussions on aspects of reading, publishing and dissemination in a "digital age," It was frequently tough going. I was favorably impressed by Roger Chartier's piece on "Readers and Reading in the Electronic Age," where he notes other people's "obsessive theme" of the "disappearance of the book." I quoted one key paragraph:

> Even without imagining this still hypothetical future, one may wonder whether the electronic book in its current form will be able to attract or produce readers. The long history of reading clearly shows that revolutions in the order of practice always lag behind, and are often slower than, revolutions in technology. New ways of reading did not follow immediately from the invention of printing. Similarly, the intellectual categories which we associate with the world of texts will endure with the new forms of book. It might be useful to remember that, after the invention of the codes and the disappearance of the scroll, the 'book'—here meant as ordered discourse—often corresponded to the textual matter previously contained in a scroll.

Chartier was also concerned that the "electronic revolution" could deepen inequalities (third-world nations have ridiculously cheap paperback books but typically not great computing infrastructures). Some commenters dismissed any doubts about the inevitable revolution. One lamented the loss of mental depth but seemed to think that, too, was inevitable.

The second paper, in translation, began: "The subject of this piece is the metaphysics of the book. I shall look at the way in which the Web frees it from our inadequate conception of it." The paper and my notes on it both gave me a headache.

## October 2002: Number 27

My favorite piece in a scattered issue was a reprint from my "disContent" column in *eContent Magazine*—remember when I was reprinting select "disContent" columns?—about *Cubed: Media About Media About Media*. The idea was to take all the metamedia of the day and go one step further: ThriceRemoved.com would focus entirely on metamedia. It would "set the record straight about journalism about journalism" and it would, of course, combine a print magazine, a website with some free content to entice you to subscribe, and a blog.

There were five factual paragraphs in an otherwise mythical "press release" and attachment, still true today (although some metamedia, such as *Brill's Content*, has disappeared):

> First, there were media—discussions and examples of real life but at one remove. In recent decades, there have grown to be so many media that life is something you do in between exposures.
>
> Media about media—metamedia—add another layer of separation: You can think of them as media squared. These aren't new, to be sure. What newspaper doesn't include columns and reviews on other media—TV, books, movies, the Web?
>
> Even media *wholly* about media go back decades. Consider *Variety* and *Broadcasting & Cable*, *Editor & Publisher*, *Publisher's Weekly*, just to name a few. *Columbia Journalism Review* and *St. Louis Journalism Review* are long-standing examples of critical media about media (or journalism about journalism).
>
> The trend has grown in recent years—and there are new media to have media about media. *American Journalism Review*, *Publish!*, *Brill's Content,* even *EContent*; on the radio, "On the Media"; on TV, "Talk Soup" and 'serious' discussions of media elsewhere.
>
> The Internet? Where to begin? In addition to the Web versions of print metamedia, there's MediaWeek.com, the *Online Journalism Review*, and of course the stunning success of [Inside].com. A whole new medium, Weblogging, has encouraged many more media about media, most noticeably Jim Romanesko's MediaNews.

## July 2003: Number 37

"Why make records when you can make enemies?" That was the headline over a Copyright Perspective focused on RIAA—and some things haven't changed much in five years.

> Why does all of this matter to libraries and librarians? Because the ultimate goal of Big Media appears to be absolute pay per use with absolute control over all uses of "their" songs, books, movies, etc.—and if you think libraries will be exempted, think again…

Nor are proposed "remedies" that could damage your computer even if you've done nothing wrong anything new. I started with some quotes from *Media Life* on RIAA tactics.

> Meanwhile, "With the help of unnamed technology firms, music companies are quietly looking into ways to interfere with pirates' ability to download files, including such guerilla tactics as knocking potential downloaders offline and even messing with their hard drives."…
>
> Set aside that word "pirates," a largely incorrect term that deliberately biases reader opinion. As the next sentence notes, "That some of these tactics may be illegal to carry out has so far not deterred record labels." In which case, the record labels are setting themselves up for to be on the other side of lawsuits. How about this one: "One program would scan computer hard drives for pirated music files and automatically delete them."
>
> Repeat after me: *There is no way to determine that an MP3 file on a computer has been "pirated*." I know of no reliable way to differentiate between the MP3 files on my PC (all legitimately ripped from CDs that I own, *none* of which are available for sharing) with MP3 files that Scurvy Jack copied from Peg-Leg Pete's MP3-o-rama site… Deleting any of those files would be an act of destructive cracking and should be subject to substantial civil or criminal penalties. The article mentions wiretap laws; there must be others that apply.

I'll stand by the italicized statement above. I have more than 1,000 MP3 files on my PC. Almost all of them are of commercial music. Other than the high bitrate (I rip at 320Kbps), there is *simply no way* to differentiate these files from illegal downloads. None.

## March 2004: Issue 47

This issue had one of those new continuing features that somehow never continued: THE WAY WE'RE WIRED. It started with eloquent reader feedback regarding my comment that Amazon harms local booksellers even while I was praising NetFlix, which could be seen as harming local video rental stores. "Isn't it at best inconsistent and at worst hypocritical to criticize Amazon but support NetFlix?"

> I haven't said this in quite a while, and perhaps never in a sufficiently straightforward manner.:
>
> - If you have a locally owned video/DVD store in your neighborhood that stocks the movies you want to rent, and you find that store an agreeable place to do business, you should *certainly* favor that store over NetFlix.
> - Conversely, if there are no locally owned bookstores in your area, or you are repelled by the local bookstores, then you should evaluate chain stores and internet bookstores to see which ones suit you best…
>
> For us, the choice was easy. I'd had a six-month trial NetFlix membership… but we were renting most of our DVDs at a good local video/DVD store. About the time the NetFlix freebie was going to expire, the local store disappeared, thanks to rent gouging by the mall owner. That left two choices: local Blockbuster franchises or NetFlix. I don't care for Blockbuster, for a variety of reasons. I like NetFlix.

That's my situation. Yours may differ. For some people, a combination makes most sense: A good local store for mainstream DVDs, a minimum-level subscription to NetFlix for the stuff the local store doesn't handle. There's the library too, to be sure…

I believe that local video stores have disappeared to a much larger extent than local bookstores. I believe—without much proof—that Barnes & Noble and Borders, while certainly not as "local" as good independent stores, are reasonable alternatives when no good local store is available, where I have no such belief when it comes to Blockbuster and Hollywood Video. But yes, maybe I am inconsistent, possibly even hypocritical.

### *The General Case and the Way We're Wired*

This *mea culpa* appears under the "Way We're Wired" flag because it's an example of legitimate differences in preferences and habits.

Set aside for the moment local tax revenue issues. Those can be solved… Fact is, some people simply don't want to deal with certain businesses and have preferred ways of buying that send them to the internet, or to chain stores, or wherever.

I don't have a problem with that. If that's your preference, that's the way it is.

For some of us, maybe most, it depends on the kind of product and the nature of the local stores…

There are local bookstores that drive away customers. I've read a science fiction/fantasy magazine editor's comments on being informed that her local bookstore didn't sell "that kind of book" and wouldn't special order such trash. There's nothing wrong with a store's stock reflecting the owner's preferences and with the staff revealing their tastes—but there's also nothing wrong with customers going elsewhere…

Problems arise when you do your browsing and sampling at the retail store, then buy on the internet to save a buck or two. The extreme case comes with goods such as high-end audio, where you may be using a significant amount of staff time to explore choices. I think there's an ethical issue involved here, and it's a direct way to undermine local business…

I hope Amazon doesn't become the only game in town. For that matter, I hope NetFlix doesn't become the only game in town. I don't think either one is likely. Diversity in the marketplace is almost always a good thing.

Amazon has not become the only game in town. NetFlix has not become the only game in town. I do buy from Amazon—but rarely books (Mountain View does have bookstores).

## December 2004: Issue 57

From INTERESTING & PECULIAR PRODUCTS:

Here it is, as promoted in the September 2004 *Computer Shopper*: wearable television! NHJ's $219 VTV-101 TV-Wristwatch, with a 1.5" color screen (0.9" high and 1.2" wide, I presume). "It's rated to run on an internal battery for *about an hour* and an external battery pack of four AA batteries for 3 hours." [Emphasis added.] They don't provide overall dimensions, but it looks to be a little over two inches in each dimension and more than half an inch thick. In other words, one *seriously* geeky watch. Wonder how long it actually functions as a watch after you've watched TV for an hour? It's designed for the Japanese market, which explains a lot.

## October 2005: Issue 67

Here's an oddity:

"Art finds a mobile home" in the June 2005 *EContent* is mostly an interesting story about using mobile technology to "bring art to the masses and to provide artists with new outlets and creative forms." It features the Digital Museum of Modern Art, an entirely virtual museum…

Things get weirder early on. DMOMA's founder calls mobile "a perfect medium for art" because "it allows users to bypass elite gallery systems and experience art on their terms"—and goes on, "All art can be reduced to a sequence of binary bits—zeroes and ones in endless succession." One of the artists exults, "There's no one and nothing between the user and the art. There is no distraction." That 2" screen? Not a distraction. Reducing, say, *Guernica* or any Rodin sculpture to "zeroes and ones in endless succession" viewed on a tiny screen? You've eliminated elitism and any gap between the user and the art.

Douglas Rushkoff provides the final word: "Because *art is no longer a physical thing*, it has a disposable quality to it. When something is temporary, artists are going to have to create more of it." Rushkoff's a communications professor, so his declaration that physical art is already over, kaput, finis is presumably not just sloppy communication.

Rushkoff, having written off all physical art, thinks people won't go for $35 for *one* piece of art on the phone, but might pay "$135 to subscribe to two months of images from certain artists." Wow: Another $67.50 a month for some transient art on a 2" screen. Makes museum memberships look really cheap.

I'm not saying you can't have successful art experiences on the tiny screen. I am saying that reducing "all art" to a bunch of bits and bytes and proclaiming the end of art as a physical thing is…well, I suppose "philistine" isn't politically correct. One comment along the way is just plain strange: "Much art is in galleries or in private collections. Mobile makes it possible for anyone to see art. It's no longer a privilege for the few." But the art in private collections *won't be available* on mobile phones—or has "art" become some interchangeable thing like sand or water?

## May 2006: Issue 77

The lead essay was BOOKS, BLOGS & STYLE, and I think it stands up well—well enough that instead of using a page here, I'll suggest you **go read it**.

## March 2007: Issue 87

Here's a piece from MY BACK PAGES that really wasn't snarky enough to fit there: "Free and easy—and legit":

> My freebie subscription to *Business 2.0* is almost over. I'm delighted, particularly given the "web page" redesign of the magazine—but mostly given its consistent "the only ethical issues in business are whether you make money and whether you are provably criminal." Spam link pages? Hey, you can make a buck. Domain ranching? Wow, profit ahead. But that's a topic for a blog post one of these days.
>
> This one's a "What's Cool/Playing the Angles" one-pager, "Free and easy," in the December 2006 issue (before the big redesign, when the magazine still had serif type and justified margins, and lacked hot buttons on every page). It discusses Mat LaClear's website, selling four books on real estate. The first one, *No Holds Barred: Mugging Tactics for Today's Real Estate Agent*, costs $50 and made LaClear $15,000 in its first month. "But here's the rub: He didn't even write it." It's a slightly revised version of *Closing the Sale* by J.C. Aspley, a 1925 title that apparently didn't have its copyright renewed. (Anything before 1923 is automatically fair game in the U.S.)
>
> LaClear finds books in the public domain that might be salable today, possibly with a little reworking. He scans the book, does a bit of updating, adds a new title and puts it out as a new book under his name. He's not the only one reselling public domain material; the story mentions several others (and Dover used to publish a number of books entirely based on public domain material). The key to LaClear's success appears to be targeted email lists—that, and the apparent willingness of would-be real estate hotshots to pay big bucks for these how-to books.
>
> Here's the final paragraph:
>
>> Easy—but what about sleazy? The J.C. Aspleys of the world might bristle at content resellers basically plagiarizing their work for profit. The resellers don't see it that way, of course. "Is it fair to mankind that good, helpful manuals are lost forever?" LaClear asks. "I view myself as a recycler."
>
> So do I—although LaClear really should include a back-of-title-page note that "Portions of this book originally appeared in different form as *Closing the Sale* by J.C. Aspley." Claiming authorship for the "new" book if he's just rewritten a couple of anecdotes *is* sleazy, and would be entirely unacceptable for scholarly work.
>
> But reusing public domain material in new works is neither illegal nor unethical, and I wouldn't call it sleazy if appropriate credit is given. The public domain's *supposed* to generate new content and new uses for old content.

## December 2007: Issue 97

Let's end with a segment of the one and only TRENDS & QUICK TAKES PERSPECTIVE:

> Back to Nicholas Carr's *Rough type* for a cute little post on that always-right guru, Ray Kurzweil:
>
>> I was flipping through the new issue of *The Atlantic* today when I came across this announcement from Ray Kurzweil: "The means of creativity have now been democratized. For example, anyone with an inexpensive high-definition video camera and a personal computer can create a high-quality, full-length motion picture." Yep. Just as the invention of the pencil made it possible for anyone to write a high-quality novel. And just as that power saw down in my cellar makes it possible for me to build a high-quality chest of drawers.
>
> The *tools* have been democratized: "Inexpensive high-definition video camera" is no longer an oxymoron (depending on your definition of "inexpensive") and today's under-$1,000 desktops have more than enough processing power (and inexpensive software to use it) to do nonlinear video editing that would have required an AVID or a high-end graphics workstation a few years ago.
>
> None of which democratizes creativity. I'm unlikely to write a great novel, no matter how well Word works—and switching to different software on a faster PC wouldn't help. I'm unlikely to compose great music even if I buy the appropriate software and hardware. I'm *extremely* unlikely to make a great movie.
>
> It's not just creativity. Most creative works involve effort as well as talent. Better tools may lessen portions of the effort, but it doesn't go away. A good flick involves a host of different talents; unlike most fiction (and nonfiction), it's almost always a deeply collaborative proposition involving not only different kinds of brainwork but a fair amount of brawn as well. All the nonlinear editing software in the world isn't going to give one of the friends I could con into doing a home movie the acting talent of Reese Witherspoon or Mary Kay Place or the composing talent of Randy Newman.

Or, in one brief sentence: This stuff is hard.

# Masthead

*Cites & Insights: Crawford at Large,* Volume 8, Number 11, Whole Issue 108, ISSN 1534-0937, a journal of libraries, policy, technology and media, is written and produced by Walt Crawford, Director and Managing Editor of the PALINET Leadership Network.

*Cites & Insights* is sponsored by YBP Library Services, http://www.ybp.com.

Opinions herein may not represent those of PALINET or YBP Library Services.

Comments should be sent to waltcrawford@gmail.com. 

URL: citesandinsights.info/civ8i11.pdf