Cites & Insights

Crawford at Large
Libraries • Policy • Technology • Media

Perspective

Discovering Books: The OCA/GBS Saga Continues

The short version could be one paragraph. New members continue to join the Open Content Alliance, with affiliated projects such as Alouette, involving 27 major Canadian academic research libraries, and a group of committees have formed to plan OCA's future. The Google Library Project keeps scanning, the lawsuits haven't been settled, Google continues to be more opaque than seems necessary—and Google Book Search generates lots of articles and discussions.

Open Content Alliance

Jeffrey Young wrote about OCA in "Scribes of the digital era," *Chronicle of Higher Education* 52:21 (January 27, 2006). Young says "many see" the project as "primarily a response to the controversial book-scanning project led by Google," and that's unfortunate.

"Although the Open Content Alliance has pledged not to scan copyrighted works without permission, thereby avoiding that thorny legal issue, the project could do as much to shake up the library world as Google's effort has." I would question whether either project does all that much to "shake up the library world," but maybe I'm dense. Young seems to suggest this is the first time libraries have worked together toward digital archives and quotes Brewster Kahle on it being "a vision for an open library." As of January 27, some 34 libraries had joined OCA.

It's an interesting article that includes a description of the Scribe, the document scanner used in OCA, which involves manual page turning by employees who can "scan about 500 pages per hour."

The article says online users will be able to order bound reproductions of OCA books "by paying a small fee to a company that does the printing and binding." I don't see such a function yet at openlibrary.org, the site at which sample OCA books can be read, but the "book" about Open Library indicates this as a possibility.

The story reasonably contrasts OCA's open model with Google's continuing opacity about GBS and the Google Library Project. Google's comment on OCA, as cited in the article: "We welcome efforts to make information accessible to the world, The OCA is focused on collecting out-of-copyright works which constitute a minority of the world's books—a valuable minority, but certainly not complete."

Inside This Issue

Next steps

OCA has published its 2006 work agenda at www.opencontentalliance.org/nextsteps.html.

> The OCA will initially concentrate on digitally reformatted monographs and serials which represent diverse times, regions and subjects which are in the public domain or available under a Creative Commons license. In other words, the OCA is initially interested in the broad range of digitized documents that are in our libraries and archives.

> For an October 2006 event, we would like to focus on materials that reflect the history, people, culture, and ecology of North America. This decision is in part a practical one. It establishes essential priorities for the OCA while emphasizing collection depth as a means of encouraging the development of value-added services. It also reflects the general orientation of the initial collec-

tions that have been offered to the OCA (at this stage, OCA is not harvesting metadata).

The coalition has also established six working groups to "advise on key operational issues and help establish essential policies and practices:" Metadata and collaborative collection development, digital preservation, contribution, book format, scanning protocol, and data transfer protocol. The groups are chaired by key people at the California Digital Library, Internet Archive, RLG and OCLC; each group is supported by an RLG staff member with appropriate expertise (for example, Robin Dale serves as RLG program officer support for the digital preservation working group).

Jim Michalko of RLG commented on the 2006 agenda in a January 12, 2006 *hangingtogether.org* post. He notes "a few key things" in early organizational steps "that please me":

> keeping the OCA a project of the Internet Archive sidesteps the kinks that long, premature conversations about governance would engender

> declaring a collection focus—Americana, specifically North Americana—provides an essential filter for quick progress and priority setting

> stating a target of October 2006 to unveil a significant digital collection allows all the contributors to focus their efforts.

Related projects and new participants

What about the Million Book Project? The stated goal of the project was to scan one million books by 2005. That goal was clearly not reached. Notably, 10,532 scanned books from this project were available at the Internet Archive two years ago—and the number has increased to 10,556 as of March 22, 2006, despite Brewster Kahle's assurance in December 2004 that "tens of thousands" were on the way. According to MBP's FAQ, some 600,000 books have been scanned (primarily in India), but these are not all available online—and, indeed, I can't find any indication of how many *are* online.

Note this assertion at the Indian center: "The technological advances today make it possible to think in terms of storing all the knowledge of the human race in digital form by the year 2008." I find that a trifle optimistic. It appears that the project is becoming affiliated with OCA, to some extent. It clearly can't be accused of being Anglocentric: Of the 600,000 books scanned, roughly 135,000 are in English.

A December 29, 2005 note at *CBC Arts* (www.cbc.ca) adds a larger Canadian perspective to the early involvement of the University of Toronto in

OCA: 27 major Canadian academic research libraries have joined the Alouette Canada project, a digitization alliance with a substantial scope. According to the release, Alouette Canada "is working with" OCA and also focuses on works already in the public domain.

Google Book Search: Brief Items

Some smaller items about GBS and the Google Library Project (GLP), in chronological order:

➢ A December 12, 2005 *Library Journal* item notes the difficulty of finding the "Find it in a library" link, mostly because it *only* appears on GLP books, not the "much larger (for now)" collection from publishers. That's an issue Google needs to address; in my informal testing, the library link wasn't showing even for some items clearly in the public domain. (When James Jacobs at *diglet* asked Google about this, he received a reply stating the facts, with no explanation. As Jacobs notes, following Google's reasoning, GLP books should *not* have links to online booksellers.)

➢ Mary Sue Coleman, President of the University of Michigan, spoke on "Google, the Khmer Rouge and the public good" to AAP's Professional/Scholarly Publishing Division on February 6, 2006. She strongly defends GLP and Michigan's role, explaining why Michigan considers it "a legal, ethical, and noble endeavor that will transform our society." I won't go into details of the talk, which is readily available online, but would note that Coleman stresses the *preservation* aspect of GLP—and that turns out to be a tricky topic (see below). Apart from that issue, I believe Coleman gets it right.

➢ Siva Vaidhyanathan seems to have moved from an argument that GLP is a bad test case for fair use to a more general condemnation of Google. He now denounces GLP on several grounds—and concludes, apparently, that he knows more about librarianship than the directors of the Michigan, Stanford, Oxford, Harvard, and New York Public Libraries. He calls Coleman's speech "disingenuous," says that GBS offers "stunningly bad results" and offers libraries the arcane advice "Don't throw away that card catalog just yet." He calls the deal with Google "horrible," and says "it is stupid and counterproductive" for librarians

to "sign over control to an unaccountable private entity." He says "libraries that are giving away the treasure have abrogated their responsibility to defend the very values that librarianship supports." Vaidhyanathan claims to be pro-librarian/pro-library. Gale Norton claims to be an environmentalist. I, for one, was not aware that librarians were "signing over control" through participation in GLP or that librarians were "giving away the treasure" by *lending copies of books* (which is, after all, one of the things libraries do). Michael Madison at *madisonian.net* has been arguing some of these issues with Vaidhyanathan; Madison doesn't seem to think we need to "stop Google to save librarians," and I agree. Apparently, one of Vaidhyanathan's arguments is that he had trouble finding Cory Doctorow's *Down and out in the Magic Kingdom* using GBS—but searching for "science fiction magic kingdom" *in Google* yields the book right away. The problem here is that Doctorow's novel apparently isn't *in* GBS—so you can't find it there, although it's *readily* available through Google itself.

➢ James Jacobs posted something much more significant at *diglet* on February 16, 2006, in "Thoughts on Google Book Search," after hearing Daniel Clancy, engineering director for GBS, speak at Stanford: "Clancy mentioned that Google was *not* going for archival quality (indeed *could* not) in their scans and were ok with skipped pages, missing content and less than perfect OCR—he mentioned that the OCR process *averaged* one word error per page of every book scanned! The key point that I took away from this is that Google book project *is not* an alternative to library/archive/archival/preservation scans. Libraries will still have an important role to play (as we already know!) because a certain percentage of the digitized content owned by StanMichOxYork will be basically unusable as archival, preservation-level digital content. Google's ok with that, but libraries shouldn't be!" For a book search engine, one word error per page isn't bad (that's roughly 99.7% perfect OCR)—but it appears that Mary Sue Coleman may have received a scrambled message about preservation.

➢ Cory Doctorow thinks publishers "should send fruit-baskets to Google" and explains why in a February 14, 2006 essay at *boing boing*. I disagree with Doctorow on huge chunks of his argument (print books are going away, people now get all their info online, yada yada), but he makes excellent points on some of publisher and author complaints against Google, specifically the idea that because Google intends to make money (indirectly) from GBS, authors and publishers should get a cut of the action. "No one comes after carpenters for a slice of bookshelf revenue. Ford doesn't get money from Nokia every time they sell a cigarette-lighter phone-charger. The mere fact of making money isn't enough to warrant owing something to the company that made the product you're improving." It's a *long* essay, particularly for *boing boing*—4,096 words, the equivalent of more than five *C&I* pages. (Commenting on Doctorow's essay, Vaidhyanathan says "the case law on fair use is totally hostile to Google," despite Doctorow's citation of case law that *favors* Google. Lawyer Jonathan Band, cited below, also believes that there's significant case law favoring Google. Apparently, Siva Vaidhyanathan is not only a better librarian than five major library directors, he's a better copyright lawyer than Jonathan Band or others who believe Google has a good case—since he says "totally hostile," it must be overwhelming. I'm impressed by the multifaceted genius and authority of Prof. Vaidhyanathan!)

➢ A February 23, 2006 *Chronicle of Higher Education* piece by Andrea L. Foster notes Google's new "fact-checking brigade" to cope with "misperceptions" about GBS. One such misperception is Susan Cheever's *Newsday* assault on Google. Among other things, Cheever says, "The amount of words that constitute fair use varies according to court case. At present, it is 400 words." As any librarian should know, *that's nonsense*. The U.S. Copyright Office fact sheet does not provide a word limit. Even the conservative guidelines from the office suggest "1,000 words or 10 percent of a work of prose, whichever is less" for republication—and those are guidelines, not legal findings.

> Rob Capriccioso wrote "Google's not-so-simple side" on February 27, 2006 at *Inside higher ed* (www.insidehighered.com). He reports on a "lively discussion" at the American Enterprise Institute-Brookings Joint Center for Regulatory Studies. One audience member made the claim that "it would be relatively easy…to quickly piece together snippets of…books until entire chapters or texts were available online," a claim that's almost certainly nonsense. Edward Timberlake, "who said he works at the U.S. Copyright Office," made a startling statement about the copies of scans that Google returns to the owning libraries: "He said that the libraries are doing 'a lot of stuff' with those electronic versions that authors and publishers don't believe they have permission to do." But authors and publishers chose *not* to include the libraries in their suits against Google, and there is *absolutely no indication* that any library involved plans to do anything other than use the scans as dark archives. Capriccioso doesn't cite any *example* from Timberlake of this "stuff" libraries are doing.

Google Book Search: Longer items

Congressional Research Service

Robin Jeweler of the Congressional Research Service prepared "The Google Book Search Project: Is online indexing a fair use under copyright law?", issued December 28, 2005 (Order Code RS22356, available at fpc.state.gov/documents/organization/59028.pdf). The six-page report notes the situation and that, "Once again, new technology and traditional principles of copyright law appear to be in conflict."

> Because of the unique facts and issues presented, there is scant legal precedent to legitimize Google's claim that its project is protected by copyright law's fair use exception to liability for infringement. Thus, questions presented may be ones of first impression for the courts.

Jeweler concludes that Google's "opt out" option "contributes to the content holders' claim that Google is engaged in massive copyright infringement." Summarizing the positions, Jeweler says plaintiffs consider Google's project strictly commercial "because it 'pays' for the libraries' collections by delivering digital copies back to them" and because Google will gain advertising revenues. Google "essentially contends that its opt out program negates any infringement liability"

and that, in any case, the activity is fair use, citing Kelly v. Arriba Soft.

Jeweler notes that fair use is not strictly a matter of evaluating the four factors encoded in law; "Because fair use is an 'equitable rule of reason' to be applied in light of the overall purposes of the Copyright Act, other relevant factors may also be considered." Without attempting to predict how courts would rule, the CRS report offers some observations on the issues at hand. A few examples:

> With respect to the first factor, the purpose and character of use, the searching and indexing goal appears to be a highly transformative use of the copied text. There is little question that indexing basic information about any book alone, absent copying, would not constitute copyright infringement. While displaying "snippets" of text is closer to infringing activity, the prospective display, as described by Google, does not appear to usurp or negate the value of the underlying work.

> The second factor is the nature of the copyrighted work. Digitizing the collections of the named libraries will encompass both factual and creative works, the latter being entitled to the highest level of copyright protection. How the court views the third factor—amount of the portion used—will be significant. In order to create its megadatabase, Google will scan the entire copyrighted work, a major consideration weighing against fair use. But it intends to display, i.e., use, at any given time, only brief excerpts of the searchable text. Hence, is the digital reproduction incidental to an otherwise fair use or is it impermissibly infringing?

> Finally, what will be the Library Project's effect on the potential market for or value of the copyrighted works? Here, Google makes a strong argument that its indexing and text searching capability has the potential to greatly enhance the market for sales for books that might otherwise be relegated to obscurity. Its "sampling" of text permits members of the public to determine whether they wish to acquire the book.

Jeweler notes publishers' claim that copyright owners routinely receive license fees for authorized sampling (but not, as far as I know, for indexing). There's the speculative claim—publisher *could* potentially participate in, and derive revenue from, a similar project. And, of course, publishers "expressed concern" that the library copy "may facilitate piracy and/or additional unauthorized uses"—although publishers didn't sue the libraries.

How about case law? "Google asserts that Kelly v. Arriba Soft Corp. supports its claim of fair use, and in many respects it does." Google's snippets represent "far more limited reproduction and display" than Arriba Soft's thumbnail images of full-sized pictures. A

distinction is that the images in question were voluntarily uploaded to the internet.

There's more. *Sony Corporation of America v. Universal City Studios*—the Betamax case—held that, in some cases, apparently infringing activity that facilitates an arguably legitimate use is fair use. Other cases have failed to expand that category—but neither have they overruled it. The report concludes:

> How the court (or courts) that consider this case define the issues presented will ultimately determine whether the suit against Google sets an important precedent in copyright law. Viewed expansively, the court may find that copying to promote online searching and indexing of literary works is a fair use. To many observers, such a holding could be the jurisprudential equivalent of *Sony's* sanctioning of "time shifting." If the court adopts a more narrow view of fair use that precludes Google's digitization project, searchable literary databases are likely to evolve in a less comprehensive manner but with the input and control of rights holders who view them as desirable and participate accordingly.

Jonathan Band via ALA OITP and Plagiary

Jonathan Band continues to write some of the most lucid analyses of GLP. *The Google Library Project: The copyright debate*, issued in January 2006, is available as an OITP Technology Brief from ALA at www.ala.org/ala/washoff/oitp/googlepaprfnl.pdf. A related analysis appears in the new ejournal *Plagiary* (www.plagiary.org) as "The Google Library Project: Both sides of the story."

Both sixteen-page publications provide detailed discussion of the issues at play. Unlike far too many commentators, Band is very clear about the limited visibility of copyright works: "This is a critical fact that bears repeating: for books still under copyright, users will be able to see only a few sentences on either side of the search term—what Google calls a 'snippet' of text… Indeed, users will never even see a single page of an in-copyright book scanned as part of the Library Project." Here's one I hadn't realized: "Google will not display any snippets for certain reference works, such as dictionaries, where the display of even snippets could harm the market for the work."

Band finds *Kelly v. Arriba Soft* applicable, and goes a little further than the CRS report: "[I]t is hard to imagine how the Library Project could actually harm the market for books, given the limited amount of text a user will be able to view… Moreover, the Library Project may actually benefit the market for books…"

Publishers claim Google's storage of the full text of each book makes it different from Arriba's storage of compressed low-rez versions of images. Band: "This seems to be a distinction without a difference, because Arriba had to make a high resolution copy before compressing it." Publishers also attempt to deny the applicability of Kelly because it involved the copying of digital images already on the internet (thus providing an implied license to copy), while Google is digitizing analog works.

> Google has three possible responses to this argument. One, the *Kelly* decision makes no reference to an implied license, nor has any other copyright decision relating to the Internet. Two, this argument suggests that works uploaded onto the Internet are entitled to less protection than analog works. This runs contrary to the entertainment industry's repeated assertion that copyright law applies to the Internet in precisely the same manner as it applies to the analog environment.
>
> Three, Google can argue that its opt-out feature constitutes a similar form of implied license…

As you'd expect, copyright holders have a third argument against applying *Kelly*: It was wrongly decided. Plaintiffs would much prefer that *UMG Recordings v. MP3.com* be used as precedent. But, Band says, Google will contend that *MP3.com* is easily distinguishable: Google's use is far more transformative and Google's use will not harm any likely market for the books. Band says "there is no market for licensing books for inclusion in digital indices of the sort envisioned by Google."

There's a lot more here, to be sure. I **strongly recommend** reading one or both of Band's pieces. He has something to say about Siva Vaidhyanathan (quoting from the *Plagiary* article, where there's a direct endnote to Vaidhyanathan):

> While in theory it might be preferable from a societal point of view for the Library Project to be conducted by libraries rather than a private corporation, libraries simply do not have the resources to do so. Thus, as practical matter, only a large search engine such as Google has both the resources and the incentive to perform this activity.

Band concludes "A court correctly applying the fair use doctrine as an equitable rule of reason should permit Google's Library Project to proceed."

EContent and Online

Jessica Dye's "Scanning the stacks" appears in the January/February 2006 *EContent*; the March/April 2006 *Online* includes a ten-page cluster of four brief articles on GBS. Both are worth reading. Jessica Dye offers a reasonable quick overview of the situation, perhaps favoring anti-Google voices somewhat.

The cluster in *Online* is curious. Marydee Ojala begins with a clear commentary on how GBS actually works, at least in its current form—and hopes that searchability improves as it evolves. K. Matthew Dames argues that library organizations should support GBS—but says that "the library community's only public comments on Google Book Search come from an ALA president who seems more concerned with the possibility that his copyright could be 'flaunted' than the possibilities that someone could find, use, or buy his work." I don't understand this: *Cites & Insights* is most certainly part of the library community, as are many blogs and periodicals that have had *very* public statements in favor of GBS. Or does Dames only consider statements by officers of library organizations? David Dillard, speaking from a reference librarian's perspective, thinks GBS can be very helpful when looking for books with relatively obscure content, offers some examples, and concludes that "revenue brought in by books should invariably increase as more people learn of books containing answers to their information needs." As with other librarians (whose opinions I've read) who have actually looked at GBS and its potential, Dillard expects it to be a good thing both for book publishing and for libraries.

Then there's Michael A. Banks and "An author looks at Google Book Search." It's the same-old, same-old. The illustrations show *entirely* books provided through the Google Publisher Project, showing no snippets at all. Banks claims GBS "can actually discourage some users from buying books" because it "displays the very information being sought" in certain kinds of nonfiction books. "Having seen the information, there's little chance the searcher will buy the books." That might be true, *if* snippets were more than a sentence or two and *if* GBS didn't suppress snippets in reference works. He speaks of "pillaged" books that are "intellectual property with value, created by people who anticipate being paid for the time, effort, and expense that go into them." Great, except for the preface: "[M]any, many readers buy reference, tutorial, and how-to books to get at specific information. Now they can go to Google Book Search and get the information for nothing." Since that's *simply not true*, the rest does not follow.

Other Google Cases

While the Google Library Project suits have not yet been heard in court, other cases have been. Perfect 10 won a lawsuit regarding thumbnail images; counsel for plaintiffs in the GLP suits claimed this finding was bad news for Google's stance on GLP, while Google and EFF didn't see any precedential similarity.

Blake Field sued Google for caching an article Field had posted on his website; a Nevada district court ruled against Field, saying he had "attempted to manufacture a claim for copyright infringement against Google in hopes of making money from Google's standard practice." The court granted summary judgment on four bases: Since Field did not allege that the Googlebot's initial copy was an infringement, using the cache could not be considered direct infringement; Field didn't opt out (there was no "no archive" metatag and there *was* an explicit "allow all" robot.txt header); Google's cache is fair use; and that cache qualifies as a DMCA "safe harbor." EFF's Fred von Lohmann says the decision is "replete with interesting findings that could have important consequences for the search engine industry, the Internet Archive, the Google Library Project lawsuit, RSS republishing, and a host of other online activities."

Another district court—this one the Eastern District of Pennsylvania—rejected a civil complaint (for copyright infringement and other activities) against Google by Gordon Roy Parker, "an online publisher of sexual seduction guides" who also offers racetrack betting tips. In this case, the complaint (filed by Parker, a former paralegal) was termed "rambling" but the judge was clear that Google's caching does not constitute infringement.

The saga will continue. OCA's benefits are clear; the alliance's choice to avoid copyright issues is cautious but clears the way for more expansive uses of material. GBS is a muddier situation, not aided by Google's lack of transparency—but there seems little doubt that GBS and the Google Library Project will serve the aims of copyright, at least as stated in the Constitution: "To promote the progress of science and useful arts." Being able to *discover* books based on obscure content within those books doesn't substitute for library catalogs and doesn't seem to have any chance of substituting for the books themselves—but it can promote progress by making it easier to find work on which to build. How can that be a bad thing?

The Library Stuff

You may begin to see stuff other than annotated citations in THE LIBRARY STUFF. While I believe everything

in *Cites & Insights* relates to libraries and librarians, I wouldn't mind including more commentary that's *directly* relevant—including comments on how libraries use new web-related tools to improve service.

I "snuck in" a few blog posts and *LISNews* stories along with the formal articles noted here in earlier issues. At this point, I don't see a useful distinction. A 3,600-word analysis by Lorcan Dempsey that appears as a blog post certainly deserves citation and comment as much as a 700-word column in *American Libraries*, at least if the content inspires me to cite it. There may still be a bright line between refereed literature and everything else, but I can no longer see any good reason to draw boundaries within the "informal" literature. If it's good, it's good.

Open J-Gate

Luke Rosenberger at *lbr* alerted me to this new e-journal portal, www.openj-gate.com, mentioned in passing in *C&I* 6:5. Open J-Gate comes from Informatics India Ltd. and proclaims itself the "portal with the largest number of e-journals." It's an offshoot of J-Gate, which is *not* a free service; J-Gate claims to index more than 14,000 journals including free access to more than 4,000. At this writing, Open J-Gate still says "3000+ Open Access Journals"—more than 1,500 of them peer-reviewed scholarly journals.

Given that the *Directory of Open Access Journals* (www.doaj.org) currently includes 2,140 journals (all of them "quality controlled scientific and scholarly journals"), with 582 searchable at article level, Open J-Gate represents an interesting complementary resource. The new player has looser standards for inclusion: thus, *C&I* is indexed in Open J-Gate but would not qualify for *DOAJ*. Open J-Gate also appears to include articles from trade publications that post some but not all of their contents on the web.

As far as I can tell, Open J-Gate does *not* provide full-text searching. It's fielded searching based on metadata. Keywords used appear to be those actually provided in metadata (it's hard to be sure; I used to add keywords for some of this e-journal's HTML pieces, but mostly don't any more). It's not a panacea, but an interesting addition.

Citations: Articles and Blog Posts

Angel, "Generational conflict? Says who? Oh, them…" *The gypsy librarian*, March 17, 2006, and Lindner, Mark, "More on generations and library literature," *…the thoughts are broken…*, March 7, 2006.

In a recent *Walt at random* post, I grumped about "gen-gen": Generational generalizations, in this case propounded by a Pew speaker. Angel does it much better, in a post that "started out as a comment and got too long" (2,580 words—yep, that is on the long side for a comment). He's partially commenting on Mark Lindner's "More on generations and library literature" (March 7, 2006, *…the thoughts are broken…*), which in turn is commenting on a *C&RL News* article about "bridging differences" across generational lines.

Here's the paragraph in the *C&RL News* piece that set Lindner off:

> *Traditionalists* are loyal employees, committed to the institutions for which they work. *Baby Boomers* are competitive and idealistic, a generation that has been able to focus on themselves. *Generations Xers*, by contrast, are skeptical and self-reliant. They have seen their parents divorce and institutions fail. Finally, *Millenials* are technologically savvy, diverse, and have been raised with a global media perspective.

Lindner's immediate response: "This is some of the most ridiculous pap that I have ever seen in print!" Followed shortly by this paragraph, which I find necessary to quote in full:

> …pretty much every librarian that I know of *any* age is committed to the institution for which they work. I, a Boomer, have lost almost all of my competitiveness thankfully. Yes, I am idealistic. Probably more so than ever in my life. But then many of my fellow students, from the ages of 23 to closer to 60, are also idealistic. By this point in our nation's history and economy most of the Boomers I know are pretty darn skeptical too. I have a degree in philosophy for cripe's sake; and that is not what made me skeptical. More the other way around. And guess what, I too saw my parents divorce. Mom twice. And I have seen my share of institutions fail. Was I supposedly sleeping while my children grew, or does this stuff just not affect those who have been labeled as being in a different "generation?" I, too, and many my age along with a large quantity of Gen Xers are technologically savvy. I am more diverse than I have ever been in my life thanks to all of the things I have been through, and I am actively working on becoming more diverse. Which when I think about it is a stupid way to state the supposed trait, but I am only responding to what was written. My children are no more diverse than I am. What a ridiculous concept. And for the global media perspective. Please, just give me a break! If we're talking about Americans here then please show me this vaunted global media perspective. Are you really claiming with any seriousness that our current media has a more global perspective than it did when I was raised? Hah! Get out a bit more. Like to another country on a

different continent. Consume some of their media and then come back and tell me American media provides a "global perspective." Been there. Done that. We fail.

I grew up skeptical (my parents helped). I've never been competitive enough for my own good, but am pretty self-reliant. I make my living through technology. That makes me…a traditionalist, since I was born in 1945. Or it makes gen-gen a crock.

Angel's comment on the *C&RL News* article: "The overall impression I got from the article is that an unknowing reader would think the generations are just fighting each other with hatchets and axes." He goes on to dissect the overgeneralizations based on his own experience and personal background. Angel provides considerably more detail; **go read it yourself**. He's skeptical, loyal when it's deserved, idealistic, technologically savvy, and "diverse" as all get out. He's chronologically a GenXer.

There's more to both posts, particularly Angel's article-length commentary. He went to school with Millenials, Boomer,s and other GenXers. "What I found is that they all bring different experiences and ideas to the table. They will all be happy to offer such ideas and share their expertise; they will even lead if given the opportunity or if they find such an opportunity. What they will not do is tolerate closemindedness and lack of insight."

I'll close with this comment, with which I agree: "The generations conflict more often than not is just a lure to confrontation. We don't need confrontation and we don't need half-baked generalizations…"

Blyberg, John, "From tech to tome: spanning the gulf," *blyberg.net*, March 8, 2006.

There have been librarians who became excellent programmers for at least three decades, probably four—and there have been programmers who learned to understand libraries for at least as long. MLS or no MLS, I'm more of a library person than I am (was) a programmer/analyst, although I've mostly earned my living as a programmer/analyst. I believe it's still true that most of RLG's programmer/analysts and nearly all of our systems designers have library degrees. So Blyberg's claimed "culture gap" between "IT" and librarians bothers me, as does the seeming suggestion that "veteran, tenured staff" are *really* on the other side of a gulf.

Never mind; despite my qualms, this is a favorable commentary. There are always misunderstandings and gaps in experience and background. Blyberg

notes some things to keep in mind in approaching people you need to work with who are on "the other side" of a real (or imagined?) culture gap. I don't think he aimed for a Top Ten, and the paragraphs aren't numbered, but there are ten paragraphs. Here are the topic sentences:

Get the dialogue started. Acknowledge the dichotomy. Make a peace offering. Make the other party comfortable. Show them you are interested. Ask them how you can help. Show them how they can help. Invite them to learn and play. Cross-train. Make plans together. Meet regularly.

It's a **good discussion, definitely worth reading**. I'd add this to "Acknowledge the dichotomy": Accept the likelihood that, if you perceive the "other side" as not understanding where you're coming from, you aren't fully up to speed on *their* concerns. Chances are, you have learning as well as teaching to do.

Dempsey, Lorcan, "Libraries, logistics and the long tail," *Lorcan Dempsey's weblog*, February 15, 2006.

Ya gotta love a blog post that starts out "[Warning: long, long, long]." So it is: 3,600 words. An article disguised as a blog post. It's **distinctly worth reading and thinking about**, which is not the same as saying you should automatically accept all of Dempsey's arguments.

Dempsey notes the discussion that libraries in aggregate contain deep and rich collections—but I'd go further: Many good libraries *by themselves* represent deep and rich collections (the "long tail," if you must), which can be augmented by broader collections when needed. His issue is whether supply and demand are handled well within a network environment, and he sees problems.

Dempsey's troubled by the relatively low use of ILL: 1.7% of overall circulation in libraries as a whole, 4.7% in academic libraries. "What this suggests is that we are not doing a very good job of aggregating supply (making it easy to find and obtain materials of interest wherever they are)." Alternatively, it means that libraries do a good job of meeting *most* demands, including relatively obscure items still held locally—that many libraries have collections that relate well to their communities. (There's a third possibility: ILL is obscure in most public libraries and some academic libraries impose barriers to its use, making it less used than it might otherwise be. I suspect all three possibilities are partly right.)

Another figure: In two research libraries across several years, roughly 20% of English-language books accounted for about 90% of circulation. That strikes me as being both predictable and natural, with few implications—but I may be wrong here as well.

There's a *lot* to think about here; I've just touched the surface. Dempsey elaborates on possible problems, how some web resources appear to solve them, considerations for libraries, and more. I don't see as much consideration of the uniquely local nature of good libraries as I'd like, but maybe I'm looking in the wrong place. I'd love to know more about actual use of my own public library's convenient aggregation methods (a substantial regional multitype union catalog that's directly suggested on catalog searches and offers fast, convenient retrieval): Does it significantly increase usage of materials from other libraries—and *is there* a substantial potential demand for such materials? Dempsey talks about "the massive expense of maintaining redundant collections," but "redundant" is a tricky word to use in a nation with thousands of independent libraries serving diverse communities (both public and academic).

"Editors' interview with Victoria Reich, director, LOCKSS program," *RLG DigiNews* 10:1, February 15, 2006.

Want to keep up with LOCKSS (Lots Of Copies Keep Stuff Safe)? (If you care about long-term access to digital journal collections, you *should* want to maintain awareness of LOCKSS.) Then **go read** this seven-page interview—and, frankly, if you're interested in digital preservation, you should be reading *RLG DigiNews* on a regular basis anyway. (No full disclosure required: I work for RLG, but have no connection with *RLG DigiNews*, which is written by staff at Cornell University Library in any case.) It's free, it's online, it's concise, and it has great stuff.

This particular great stuff updates the concepts behind LOCKSS, the state of the LOCKSS alliance (launched in 2005), and the new CLOCKSS initiative ("C" for Controlled), "designed to test the feasibility of a large, community-managed dark archive."

I won't attempt to summarize. There's a *lot* of information here, tersely presented: the bases for LOCKSS policies and procedures, relationships with publishers, how the LOCKSS polling process works, how much redundancy is needed and desired, the costs of an institutional "LOCKSS box" (one that's being evaluated is a $3,500 unit with two *terabytes* of storage—"far less powerful" than a typical desktop or laptop PC, but with loads of storage space and enough computational power to handle LOCKSS requirements), and more. **Seriously good stuff**.

Etches-Johnson, Amanda, "Shiny new toys @ your library," *blog without a library*, February 20, 2006.

I've seen a few recent posts that warm my heart, as they question the extent to which people admire shiny new toys for their newness, solutions that must be adopted whether or not a problem has been identified. This is an excellent example. Etches-Johnson plays off an *ACRLog* post and the notion that "perhaps we're implementing Web 2.0 technologies (like blogs, rss, wikis, etc.) for the sake of the technologies themselves and because they're new, cool, and we mistakenly believe that our users want them."

Etches-Johnson calls this the "really crucial question to ask themselves": "What need is this going to fulfill or what problem will this fix?" I'd suggest reversing the order. The first question should be "What problems do *we* have?" (that is, this particular library, not libraries as a whole) followed by "Does [whatever] have the potential to solve one/some of them without creating larger problems?"

She offers two examples at her library: Implementing a blog a couple years back and implementing IM reference last year. In the first case, there were "really specific needs":

➢ "We needed a better way to archive our news stories
➢ "we needed to provide more people with an easy way to add news content (without having to know html)
➢ "we needed an easy way to repurpose news content on the rest of our site

"So, yeah, not hard to guess that a blog would fulfill those needs."

As for IM reference, the librarians knew "a large number of our users worked virtually…and that most of them were on MSN," so IM reference made sense.

"There certainly is a cool-factor associated with these 'shiny new toys,' but implementing them for the sake of their 'shininess' makes no sense." She doesn't think the blog and accompanying RSS did much for her library's coolness; IM reference, though, sent it "through the roof."

Good stuff. My comments are almost as long as the post; consider this a proxy for several thoughtful

posts on several blogs about using new technology sensibly. The shiny new toys are just tools. Some make sense in some libraries and not others. None makes sense except in relation to real needs and uses for the real users. All make sense if they solve real problems within a given library.

Farkas, Meredith, "The road to hell is paved with good intentions," *Information wants to be free*, February 12, 2006.

If Farkas' book (next item) is as good as some of her two-page posts, I'll have to buy a copy (not something I do very often with the library literature!). Here she considers helping patrons, marketing library services, making libraries better—and the complications of real life. She agrees with Steven Bell that there isn't a librarian out there who doesn't want to make libraries better—and adds, "Librarians sometimes make pretty bad mistakes in the name of improving our libraries." But who's to say what's a mistake and what *does* improve libraries?

Maybe "barriers" like "no cell phones" and "no food or drink" are "non-user-centered"—or maybe not. The whole point of asking patrons not to use cell phones in libraries is so *other* patrons can study, read, whatever without some jerk shouting into a phone—and it's no great fun to work at a keyboard after the previous patron spilled a drink on it. (I agree with Farkas that libraries should have spare keyboards and mice handy, although saying keyboards cost "next to nothing" is like saying replacement DVDs cost "next to nothing"—about the same next to nothing, actually. $20 a shot adds up pretty quickly. Yes, you can get cheaper keyboards, but one that's worth using will still run $15 to $20.)

Then there are fines. Farkas makes a case for keeping fines and against the Netflix model; it's a case worth reading and considering. "Getting rid of fines may very well annoy more patrons than it would please and we shouldn't make the assumption that all of our patrons want to get rid of fines." Here's a great statement: "I worry that sometimes we are so focused on being cool that we'd risk alienating a lot of our patrons for the sake of appearing less like a 'stereotypical librarian.'"

> The key is not assuming that we know what our patrons need or want. We need to make every effort to know our patrons, rather than thinking we know them. We need to actually ask them what they want and what they think about our current services. We need to try and get

patrons involved in the decision-making process at our libraries.

Farkas goes on to note that libraries need to think about *all* the stakeholders, "not just the youngest, the loudest, or the ones with the most money." Indeed.

Farkas, Meredith, "Working on the book: Lessons learned so far," *Information wants to be free*, February 8, 2006.

Farkas is writing a book on libraries and technology. In the process, she's prepared an excellent set of ten "lessons learned so far." **Go read the post.** I'll just state the topic sentences, one heavily paraphrased, but it's the paragraphs that make them live.

> It's really important to manage the project well. Structure things the way that works best *for you* [emphasis added]. Don't be too hard on yourself. A book is a lot more than research and writing. It's a good idea not to edit anything until you've written everything. You will need to lay out the chapters...for your proposal, but don't be surprised if it all changes… If you're having trouble, talk to your editor. (Have someone to bounce your fears off of.) Books are hard to write along with a full-time job—but totally worth it. Don't underestimate yourself.

As I commented, if there's ever another edition of *First Have Something to Say*, I might ask for permission to quote the whole two-page post. While it's impossible for me not to edit as I'm writing, the kind of "editing" Farkas means—where you sit down and tear a chapter to pieces—is indeed best done after you've completed a draft. The sentence that follows her second point needs to be repeated as well: **There is no right way to write a book**.

Fenton, Eileen Gifford and Roger C. Schonfeld, "The shift away from print," *Inside Higher Ed*, December 8, 2005. insidehighered.com/views/2005/12/08/schonfeld

This article argues that transition to all-electronic form for scholarly journals, even those in the humanities and social sciences, "seems all but inevitable"—and that the shift "may endanger the viability of certain journals and even the journal literature more broadly—while not even reducing costs…"

It's an interesting treatment, although I wonder about the seeming inevitability of, say, journals in art and architecture going all-digital. If it's true that it's "become the norm" to migrate away from print even at the largest research universities, there are dangers that probably aren't being addressed very well. And, frankly, I find it specious to suggest that a "tipping

point" would mean that *any* continued print acquisition no longer makes sense.

The recommendations make sense, with caveats, and I'm not the one to state those caveats. Briefly, the authors recommend that all publishers develop a strategy to live with an all-digital journal environment; that libraries "and higher education more broadly" consider how they can support publishers that will find such a transition difficult (e.g., those scholarly societies that haven't been gouging libraries); and that libraries try to manage format transition strategically. There's another bullet, difficult to summarize. They seem to be saying some society journals just won't survive and that "the alternative may be the replacement of many of these journals with blogs, repositories, or other less formal distribution models."

Blogs?

What I *don't* see: Suggestions that libraries themselves could take over the "publishing" duties for more open-access ejournals and that such journals could replace some struggling journals without abandoning the journal's badge altogether. The quotation above seems to suggest abandoning quality control. What could require such an extreme alternative?

I'm taken aback by this statement, which doesn't even limit things to the journal literature: "The widespread migration from print to electronic seems likely to eliminate library ownership of new accessions, with licensing taking the place of purchase." Are we really at that state—where academic libraries abandon the long collection altogether?

Worth reading, but I find the piece raises more questions than it answers. Maybe that's because I'm naïve enough to believe that quality academic institutions aren't universally ready to scrap their collections. The comments are an interesting mixed lot. Steven Bell argues that journals in some disciplines aren't going to abandon print any time soon, Malcolm Compitello derides the notion of replacing scholarly journals with blogs, "dan" points out the difficulty of long-term access in a licensing environment, and Joel Bradshaw (University of Hawaii Press) offers a mild defense of "peer-reviewed blogs" as a distribution mechanism, while noting (correctly) that "unsubstantiated drivel is not limited to blog formats."

Bradshaw makes a good point—but a "blog" consisting of refereed articles is an issue-per-article ejournal using a lightweight publishing system. A sensible publisher wouldn't call it a blog, even if WordPress (for example) was the underlying software.

"Interview as learning tool," *digitize everything*, February 3, 2006 (www.digiwik.org)

The writer reports on a full day of interviewing for a digitization librarian position at their library. They note some of the views—and comment on where they do and don't agree. Naturally, "digitize everything" resonated with the blogger—but the blogger also saw the point of the preservationist who said "I just don't trust digitization," since digitization is not preservation. (There *are* initiatives that can combine the two—and since I work for RLG, which has spearheaded the development of guidelines for trusted digital archives, I'm acutely aware of the initiatives and some of the difficulties.) "It isn't enough to just digitize. We should add value." The blogger asserts that digitization itself is enough added value. The blogger also disagrees with someone saying, "Users don't want everything to be online"—and in context (oral histories), the blogger may be right, but as a general rule I think that's pushing the digital viewpoint too hard.

I find the last three questions (called "difficult…[with] no definitive answer") most interesting: How should we decide what gets digitized first, how should we market digital collections, and "to outsource or not to outsource?" In context of the last, the blogger seems to remember seeing a digitization cost study "in RLG, but I haven't gone back to look yet." Yep, it's there—not only studies but a worksheet to estimate costs of such projects.

"The life expectancies of books," *Making light*, January 28, 2006 (nielsenhayden.com/making-light/archives/007181.htm)

"We talk about immortal literature, but the vast majority of books are as mortal as we are." True enough—and the writer provides ample examples. John Cleveland: The most popular poet of his era. Read any Cleveland lately? Tastes change, leaving a lot of authors alongside the highway.

"Tell me again how unjust it is that your own books are out of print?" In fact, "falling out of print is a book's natural fate." This is a long entry (six print pages) in a blog aimed mostly at writers, I believe, and it's probably worth reminding writers that most of what you write will fade away fairly rapidly.

"Consider, then, the duration of copyrights." You know this story: From 28 years (renewable to 56), to 28 (renewable to 95), to life of the author plus 50 years, now to life plus 70. "You can't exactly draw a

line, but somewhere in there, copyright stops being about directly rewarding an author for his work." It pains the writer to hear "respectable minor authors going on about how the extension of copyright to life of the author plus 70 years is a victory for the little guy. It isn't…" It's primarily a tool to support Big Media, with special emphasis on Disney.

The post goes into fascinating details. For example, although all the original Sherlock Holmes fiction is out of copyright, the estate was still "combative"—and there was a claim that certain images associated with Sherlock Holmes (the deerstalker cap, the calabash pipe) *were* under copyright, to whoever owned rights to the early Basil Rathbone Sherlock Holmes movies. That turns out to be nonsense: An early Holmes illustrator, Sidney Paget, used the cap—and William Gillette, playing Holmes onstage from 1899 to the 1930s, provided the pipe. "I don't know which studio it was that harassed the [publishing house], but they were asserting rights they manifestly didn't own."

Anyway, those are big guys. Most writers aren't. "Life of author plus 70 years does squat for your chances of being read." What it *does* do is make it difficult for someone to *restore* your work to print if there's a new wave of popularity—because it's probably an orphan work. Worse, distant heirs tend to take "jackpot" views when a publisher proposes a "nice little reprint project": "If one publisher is interested, it must mean that some other publisher would be interested as well. There could be an auction! A movie! A theme park! Woo-hoo! Pots of money!" Only, of course, there is no pot of gold; the original reprint publisher gives up and goes away.

There's more. This is about the time that really terrific anthologies of early 20th century fantasy and science fiction should be emerging—and they're not, because it's too difficult to secure the rights to the stories. Near-eternal copyright works wonders for a few big companies and a handful of heirs—and effectively assures that minor story writers and other authors will stay forgotten.

Trends & Quick Takes
Click Fraud

Charles C. Mann writes "How click fraud could swallow the internet" in the January 2006 *Wired Magazine* (www.wired.com/wired/archive/14.01/fraud_pr.html). Not only is pay-per-click advertising big business—most

of Google's $6 billion revenue and quite possibly another $2 billion for Yahoo!—but it indirectly supports quite a few web sites, at least as welcome extra revenue. (*Walt at random* now has Google AdSense; I believe I may reach the $100 minimum for payout within five or ten years.)

But sensible advertisers won't pay as much as $10 per click-through if they believe they're being gamed. Click fraud happens; experts disagree as to how much. If it's 50% and if the ad networks can't prevent that, it could undermine the whole model. If it's 10% and most of that's caught and not billed, it's probably workable. Estimates vary. One marketing research outfit claims fraudulent clicks are "as much as 29.5%" of total clicks.

The article's fascinating. Click fraud can be casual (clicking on the ads on your own site) or malicious (setting up automated click-through routines on a *competitor's* ad to drive up the competitor's costs). There's a bizarre variant, "impression fraud," where you repeatedly *load* a page with a competitor's ad but never click on it, hoping the network will drop the nonperforming ad. A few defenses are mentioned, but as you'd expect Google and Yahoo! aren't going to release all of their methods. Interesting stuff—and worth thinking about if you're planning to run ads on your site. (Not your library's site, one would hope…)

You Can Never Have Too Much Software?

A December 2005 *PC World* story details the extent to which name-brand software now comes with "extras"—whether you want them or not. Download AOL's newest IM and you'll get AOL Explorer, PlaxoHelper and some mystery programs. Want Yahoo! Messenger? You get the Yahoo! Toolbar and modified settings. Winzip brings along the Google Toolbar (but you're warned and can avoid it). Toolbars seem to be the most common extra; these programs don't appear to be adding spyware or adware.

I wonder what those 133,000+ files on my PC are really all about. Have you ever tried to figure out what's actually on your PC?

Audible: Just Overlook the Evangelism

I have nothing against Audible.com. I don't use it because it doesn't suit my current needs. It works great for lots of people and that's good. "Audible cranks it up" by Paul Keegan in the March 2006 *Business 2.0* is a little peculiar—because the guy behind Audible seems to believe it can or should *replace* reading.

He talks about how "what we now call reading" may be changing in a profound way. As a writer, he "discerned a deeper social problem…Nobody had time to read anymore." *Nobody had time to read anymore*. What a breakthrough! Not "some people," not "my kind of people," not "those people who *should have purchased* my brilliant book," but "nobody."

There's the following note: "More than 90 million Americans drive alone to work every day, their eyes occupied but their ears and minds mostly idle." Maybe we would be better off if people's minds were involved in driving—and is there much doubt that audiobooks and podcasts take *more* of your attention than typical radio or CDs? That's secondary; I'm sure Katz isn't out to increase traffic accidents.

Later in the article, Katz notes that silent reading is relatively recent. Here's an interesting statement: "People didn't want or need text for much of human history, and there was a very rich intellectual life going back to cave days." How does Katz know this? Through oral history?

I read the article in about four minutes. *Business 2.0* articles are, apparently, available via Audible, so I could also have listened to it—in about 20 minutes, most likely. Which is one of several reasons why, although Audible could indeed become a billion dollar business, Katz might be well advised to tone down the overthrow of the printed word. (I now see greater significance in those "Don't Read" posters: Maybe Audible really means that!)

Will High-Density DVD Succeed?

Ken Belson of the *New York Times* raises that question indirectly in "Fiddling with format while DVDs burn," published December 26, 2005. He notes that both Blu-ray and HD-DVD were supposed to be unveiled (that is, players and movies) at the Consumer Electronics Show. That didn't happen; the first U.S. HD-DVD player *might* show up in April or May 2006, the first Blu-ray player in June or July. In any case, "there are growing signs…that the battle for supremacy in this multibillion-dollar market may yield a hollow victory."

Maybe, maybe not. This is another one of those "packaged media are doomed once everyone gets everything online" stories, and as with most such stories it assumes a universality that's just not there. No, legal downloads *haven't* replaced CDs. They're still less than 10% of the music market—because lots of people want to *own* their music (and get liner notes and the

like). Sure, on-demand HD programming matters—but, as a Blu-ray spokesperson says, "Average folks still want to watch the movie and buy it. It's presuming a lot to think that they will replace the model they've used for decades." That's tricky: Most people didn't buy all that many VHS videocassettes, and the DVD sell-through model is less than a decade old. By and large, we rented our VHS movies. On the other hand, Tom Southwick of Starz doesn't exactly wow me with his statement:

> "What's happening in the video arena is just like what is happening in the MP3 market. Over time, there's going to be so much available with cable on-demand and the Internet that having a library of tapes that you buy or borrow will become inconvenient."

Tapes? In 2006? The first sentence may be right. What's happening in the music market is that some people are choosing to download, while others are sticking with packaged media. That's likely to be what happens with video as well. But the reporter doesn't want messy endings. He concludes (after the Blu-ray quote): "But even average folks may learn fast when they have cheaper and more convenient options." (This assumes that downloaded movies are cheaper. So far, the first plans for downloads you can keep are *more expensive* than most DVDs!)

People *like* options. Most of us like multiple options. We have different tastes and use different options for different purposes. At our house, we have a 2-movie Netflix subscription, because we only want to see most movies once—but we also buy a few movies and TV shows, where we believe we'll want to see them more often or where we want the extras. Why be stuck with a single model?

Keeping the Public in Publishing?

That's the title on a January 3, 2006 post at *Lorcan Dempsey's weblog*, partly a commentary on an odd article by John Sutherland in the *Guardian*, "Ivory towers will fall to digital land grab." Dempsey's commentary is sensible and penetrating. He notes that consumers are likely to become more aware of the issues of shifting from traditional models (books) to digital models (ebooks) and that libraries need to consider their responsibility to the cultural and scholarly record in an age where digital publishing is, if not universal, certainly important.

Dempsey's piece is worth reading. I'm not sure I can say the same for Sutherland's rant. Sutherland has convinced himself that the Google Library Project and

Open Content Alliance are both part of a rush to "propertise" the public domain—"that deposit of printed material that currently (but not for much longer, alas) you, I, and nobody own. It will be the biggest privatisation in history, and the most profitable. Once the public domain is propertised, it will remain proprietary material forever."

Say what? In the case of the Open Content Alliance, the commitments are up front. The digitized public domain materials will be *freely* downloadable—and the books that have been digitized will be just as much in the public domain after being scanned as they were beforehand. That's true of GBS, to be sure: A public domain book that Google has scanned is still a public domain book, whether Google makes the scanned version freely available or not.

There's much odd about Sutherland's essay (some of which Dempsey comments on). He says "universities are organized around their accumulated knowledge base" and goes on,

> Universities currently have ownership of their knowledge base. They distribute it free of charge. Lectures, courses, and seminars are "given" and "taken"; books are "borrowed" and "returned." The knowledge base is added to and refreshed, in the form of new books for the library and so on, but it is essentially a university-owned asset.

He believes Bill Gates' notion of a tablet device to replace textbooks means "students will be billed for learning material as they are now billed for mobile phone calls," that much of the time "they will be buying what was once public domain material," that universities will cease to be repositories of learning—a tall order for a proposed textbook replacement. Of course, I wasn't aware that most universities distributed their lectures, courses, or even library privileges "free of charge." Maybe universities in the UK are entirely free to students and others, but that's certainly not true around here, even for public universities.

Speaking of Distributing Free of Charge…

A March 16, 2006 item in the *Chronicle of Higher Education* notes that the Alliance for Lifelong Learning Inc. (AllLearn) is shutting down. What's AllLearn? A nonprofit venture begun by Oxford, Stanford and Yale in September 2000 to provide online noncredit courses. At first, courses were only offered to alumni from those universities; in 2002, AllLearn opened up to the public as a whole.

During its existence, AllLearn attracted 11,000 students from more than 70 countries—but attracting students was a persistent problem. Alumni apparently expected TV-quality lectures prepared just for the online courses; they weren't satisfied with the taped lectures they got. Another marketing issue was that courses were *noncredit*; people are more inclined to pay for online courses that can lead to a degree.

For Fiscal 2005, AllLearn had $2.5 million revenue and $3.28 million in expenses. The final courses ended in December 2005.

People Who Liked…

"Similicio.us" tries to do for websites what Netflix does for movies: That is, "people who liked X also liked Y." Some of the results are bemusing—the #1 "similar site" for *Cites & Insights* is (drum roll) en.wikipedia.org. So people who like this ejournal also like Wikipedia? Tied for second: blyberg.net and maisonbisson.com. Then come *Ariadne*, *D-Lib*, and eprints.rclis.org.

But that was on March 9. What happens fifteen days later? Blyberg.net moves up to tie Wikipedia, *D-Lib* ties maisonbisson, and infotangle.blogsome.com moves up to tie eprints and *Ariadne*. Then there's *Walt at random*, where the four-way tie for "most similar" consists of the *ALA TechSource* blog, *Feelgoodlibrarian*, *Lorcan Dempsey's weblog*, and *Information wants to be free*—and that hasn't changed. These tools are fun, if not perhaps entirely convincing.

<u>Good Stuff Perspective</u>
Journal of Electronic Publishing Returns!

Three and a half years. That's a long time in periodical publishing. It's even longer in "internet time." I suspect many of us who used to read *JEP* gave up—we assumed it was a goner. Which was a shame. In its eight volumes, *JEP* included some interesting, provocative articles. I almost added "even if I didn't agree with some of them" as a qualifier—but one of *JEP*'s strengths was its diversity of views, so it makes sense that I wouldn't agree with all the writers.

It's back—at the University of Michigan, although under different auspices. The new home page is www.hti.umich.edu/j/jep/. Web searches may take you to the old University of Michigan Press site, but that links to the new site, part of the University of Michigan University Library's Scholarly Publishing Office.

Judith Axler Turner notes the things "we were worried about" when the last issue of *JEP* appeared in August, 2002:

> Peer review and the Internet: is peer review as good for e-journals, and will tenure and promotion committees recognize it?
>
> Publishing scholarly works on the Internet: what can or should we do differently in the multimedia environment?
>
> Archiving scholarship on the Internet: can it be done and who will do it?
>
> The economics of electronic publishing: can publishers make money and can libraries keep costs down?
>
> Self publishing and preprints: will they undermine the foundations of scholarship?

As she notes, those questions haven't been answered—but the dominant theme in the new issue is Google, not that big a deal in 2002. Turner calls search engines "our own personal idiot savants, giving us data without intelligence, facts and not knowledge."

If you haven't encountered *Journal of Electronic Publishing*, this is as good a time as any. I downloaded and annotated six of the seven articles and Turner's editor's note; most of them are interesting and worth reading. (I'm sure Frank Menchaca's "Varieties of poetry publishing and aesthetics on the internet" is worthy—but I'm not the right reader.)

Subheadings are article names; comments are brief and not meant to be conclusive.

In Google we trust?

Geoffrey Bilder claims "the trust model of the Internet is almost antithetical to the trust model of academia." He notes all the crap you encounter online and says the move of academic publishing to online precipitates a "crisis of trust" and "deprecation of traditional mechanisms of ensuring the authority and reliability of published works." Really? Peer review can work as effectively for electronic-only journals as for print journals, and I see no mass fleeing from journals as trusted sources. I'm not sure I buy this paragraph in its entirety:

> Publishers and librarians have spent a turbulent decade engaged in the transformation of their respective practices. Previously, they were primarily concerned with physical media—the commissioning, production, distribution, curation, and archiving of works of print. Now, they find themselves preoccupied with the development of analogous processes for digital content. With their attention focused on the operational aspects of a move to the digital world, they have not been as aware of the transition from the trust model of the print-based scholarly world to the trust model of the Internet.

Without qualifiers such as "academic" and "journal," that's just not true. But that doesn't keep the article from being interesting and worth reading. Bilder talks about the "internet trust antipattern" and ways to avoid it. I think Slashdot is a poor example; when I've visited, the feedback mechanisms haven't allowed me to "focus on authoritative and relevant postings," at least when discussions enter areas I know. eBay's system may be a better example. Google's "trust" system is tricky, as it equates popularity with authority.

I've just touched some of the discussion here. It's interesting—but I wouldn't quite take it at face value. **Recommended with caveats.**

Why we publish JEP

This short piece from Maria Bonn, head of Michigan's Scholarly Publishing Office, tells us why the move from Michigan to Columbia University Press didn't work out and how the new home works. Bonn summarizes economic elements of the scholarly publishing crisis:

> Library budgets are flat or declining while the cost of academic publishing is increasing and being passed on to the consumers.
>
> The market for University Press books does not generate enough revenue to support the print publication of scholarly works and the Presses have not developed a business model for electronic publication that creates significant returns. In order to continue to publish, the Presses then require significant institutional subsidies. Most academic institutions do not provide such subsidies, forcing the Presses to close, to publish non-academic books to bring in enough money to continue to operate, or to increases prices further, thus narrowing their market even more.
>
> Small publication units within the academy of scholarly societies are finding it increasingly difficult to cover the costs of editorial development and print production. These units increasingly either fold entirely or sign on with large commercial publishers (in either case relieving competitive pressure on commercial academic publishing). Since these smaller publishing ventures have traditionally been the venue for scholarship that is perceived as having less economic value (notably the humanities and the "soft" social sciences), their disappearance or loss of independence threatens important platforms for part of our intellectual dialogue and cultural heritage.

Rights problems make things worse. The academic community is implementing strategies for change, including institutional repositories, open access journals, and library-based publishing. "SPO is one such

experiment in library-based publishing"—but it's not automatically an open access publisher. Bonn says, "[O]pen access is not a desirable or viable model for all content," accepting the cross-subsidization of organizational activities as a legitimate use of revenue, for example. Thus, SPO is willing to publish both OA and "toll access" journals. In the case of *JEP*, SPO is publishing it "because it can"—because the SPO infrastructure keeps the costs down, and because SPO believes *JEP* "is important to the academic and publishing communities."

Recommended if you care about electronic publishing and scholarly publishing.

New media economy: Intellectual property and cultural insurrection

Daniel M. Downes contributes the longest article (22 pages, of which 18 are text) and one of the most striking, speaking of "the emerging culture war":

> On the one hand there are those who accept the traditional bargain between creators and society (sharing information, publicity, and reputation) and on the other hand are those who seek proprietary rights (ownership of material and all accompanying rights).

Downes asserts that the "new media economy…created consumers, not citizens"—which seems to ignore more than ten million active participants (bloggers) at the very least. He says the global commercial media system is "dominated by a small number of powerful, mostly American, transnational media corporations." News Corp? Sony? EMI? American? In some segments of commercial media—academic journals and sound recordings—the majority of the major players are European. But never mind. "Mostly" may be a fair characterization.

There's a lot I could argue with here. Downes' characterization of a mediasphere wholly controlled by a few big players ignores a whole group of balancing forces that *do* characterize *new* media, as opposed to online versions of old media: Musical groups without major labels who can become known and can sell their wares (downloadable music, CD-Rs, short-run CDs) directly; *tens of thousands* of small publishers using Amazon and other routes to make books available; somewhere between 11 and 13 million active blogs…and more. More than 50 million creations carry Creative Commons licenses; many, perhaps most of these suggest that this claim is overstated:

> Thus, ownership of the means of large-scale reproduction and distribution gives institutional publishers, record companies, and other content distributors the

ability to govern which works will enter the marketplace.

That claim refers to a 2001 document; the online world *has changed* since 2001.

Downes' discussion of copyright is, shall we say, an interesting counterpart to absolutists like Jack Valenti and the RIAA. Downes asserts, without any real evidence, that copyright *does not* promote new creative work—and that, somehow, trademark protection means that *ideas* can't be reused. He states (correctly) that copyright holders are as likely to be big businesses as individual artists or authors—and uses that to conclude, "Copyright protection does not provide incentive to produce." He calls Napster the "digital equivalent of two friends meeting somewhere to trade cassette tapes of their favorite music," which is charming but misleading. And, as I expected, he says that "financial incentives are not the only motivation for artists to create"—which hardly constitutes justification for *abandoning* financial incentives.

I'm hard pressed to recommend this article, but it's useful to see a thoughtful exposition of what I'd almost consider a pure anti-copyright stance. As one who continues to seek balance, who does write for money, and who does receive royalties for books, I'm not the most receptive audience for this piece.

Recommended for what it is—but approach with caution.

What if Wal-Mart ran a library?

I'd rather read and recommend Downes than try to comment coherently on this piece by Joseph J. Esposito. I've encountered Esposito before; in the September 2004 *Cites & Insights* I attempted to comment on his *First Monday* article attempting to prove that open access would vastly *increase* the cost of scholarly publishing. Even then, I noted that "he's one of those who appears to see libraries as nothing more than article-pushers" and who "seems to think that libraries only *license* publications"—and noted that I'd given up on writing a Cheap Shot commentary on an earlier *First Monday* article by Esposito.

This one's no better. He seems to think that academic librarians spend all their time paying bills and renewing subscriptions, notes how much better Wal-Mart could do it (fewer libraries, but much bigger ones!), and asserts that "we need to bring carefully thought-out industrial processes to the management of libraries and publishing companies." He says "we can only do this if we get beyond the increasingly shrill and adversarial pronouncements now being

made by librarians and publishers alike." This article praises Wal-Mart, appears to dismiss libraries as outmoded institutions, wants "consolidation in the library sector" and loads of outsourcing, and assures us that "resistance is futile." (Inevitability: The first resort of someone whose arguments are unsound.) Somehow "Open Content" will "threaten" universities. He refers to libraries as "marvelous cost center[s]." He asserts that the absence of a high-quality reading device is "the principal obstacle to the complete dominance of electronics over print," and the tipping point is coming soon. You want shrill and adversarial pronouncements? You need look no further than Joseph J. Esposito.

Not recommended.

Google Scholar: Potentially good for users of academic information

Frederick J. Friend contributes a medium-length article (nine text pages) that's well worth reading. (If this is a briefer note, it's because it's a briefer article—and one where I think you're better off reading Friend than reading my comments.) He discusses ways that Google Scholar can be used now, ways it could improve, and ways it *should* improve with consultation and involvement from libraries and librarians.

I would note one small problem, unfortunately typical when using claimed numbers from search engines. He says "a Google search under the words "open access" revealed 598,000,000 entries." But while those *words* now show a result that is "about" 599,000,000 (although you can only see the first thousand or less), the *phrase* "open access" yields "about" 21,900,000—and nearly all of the first 100 results are either *about* open access or *examples of* open access journals. Similarly his example in Google Scholar: "1,250,000" (now about 1,730,000)—but the *phrase* yields about 91,300. Those problems don't really detract from the article, but this is a common enough error to be worth noting—at least if you believe, as I do, that most scholars can figure out how to enter phrases.

Strongly recommended.

KWIC and dirty? Human cognition and the claims of full-text searching

Another relatively long article (14 text pages), this one by Jeffrey Garrett. I don't think I'm swayed by the fact that Garrett is a librarian; it's a good, interesting article on its own merits. He encounters assertions that vast full-text resources means memorization is obsolete—

and finds them lacking. He also finds full-text searching in and of itself inadequate: "Words simply do not signify."

Garrett is not opposed to full-text searching; not at all. He says it "can be used to enormous positive effect, can in fact be essential for serious work—or it can be abused to dumb down the educational enterprise in ways no earlier generation could have ever dreamed possible." He sees that both full-text searching and *cataloging* are needed, as are librarians. He concludes:

> In this age, it is we in the library profession who have the mission to humanize the machine and make it serve us and our communities on our own terms.

Strongly recommended.

Following Up and Feedback

It's been too long since I've run feedback. My apologies for these delayed items. First, there's a typical goof, one I corrected in a blog post but not here: The proper URL for Marylaine Block's "Information literacy: Food for thought" is marylaine.com/exlibris/xlib271.html. Thanks to Bob Duncan, first to point this out.

Dan Cornwall on video on demand

> Your note on the "Akimbo Video on Demand Player" made think that if you had readers who REALLY didn't have enough video content should try the educational video-on-demand programs from the Annenberg Foundation's learner.org. No way that I know of to pipe it to your TV set, but the streaming video is pretty steady most days.

> My favorite program is one done by an old UCLA professor of mine:

> The Western Tradition (www.learner.org/resources/series58.html)

> Seems like an easy way to save $200, and the hapless viewer might learn something.

No comment required!

Dan Cornwall on Google Book Search

> I'm still working my way through the Jan 2006 issue, but I wanted to thank you for your commentary on Vaidhyanathan, particularly for the remark that fair use will go away without being defended. I have to admit that I was swayed by the argument that perhaps Google's case wasn't the one to use and that a ruling against Google would be disastrous for us all. In addi-

tion to Vaidhyanathan's article I've read several comments from otherwise rational authors who've gone berserk on the idea of Google using book excerpts but are ok with Google indexing their web pages. I'd hate for the Supreme Court decide against GBS and say "words are words"—search engines should be opt-in as well!

Having said that, I think you make the better argument that we need to defend fair use sometime and it can't hurt having a deep pockets company doing the defending.

Now that I'm a published coauthor, *I've* got no objection to having our book (www.worldcatlibraries.org/wcpa/isbn/1573563870) scanned into GBS. Can't speak for Greenwood or my coauthors.

Steven Bell on ACRLog

This letter's *really* delayed: I received it December 20, 2005!

> Thank you for mentioning ACRLog. We are off to a pretty good start, averaging about 2200 visits a day. We are certainly trying to take on the issues and readings that academic librarians should be paying some attention to—and challenging them to think about it. I think you'll agree that Barbara Fister is one heck of a blogger. I am really glad she agreed to join our blogging team.
>
> It's true the name and URL differ, but we thought it would be useful to conform to the standard being used by ALA blogs for the domain name—which is the division abbreviation followed by "blog"—hence acrlblog.org, litablog.org, etc. But ACRLblog seemed too mundane a name (although what the ACRL leadership decided on wasn't too far off). My number one choice for the name of the blog was "A Passion for Academic Librarianship" but that got shot down—too long I guess.

My reply at the time (in part): "Yes, Barbara Fister is sensational. Feel free to pass that remark along." That continues to be true, and *ACRLog* continues to be a strong blog.

Joshua Stratton on copyright and what authors deserve

Way back in May 2005, in a ©3 Perspective, I cited a comment by Josh Stratton on a post by Lawrence Lessig. Stratton said, "I don't think that authors deserve anything" and "I'd prefer to relax the laws" rather than using Creative Commons. I commented: "I'm not sure Stratton is a full 'you wrote it, it's *mine*' anti-copyright advocate, but he's close." I also noted that, if possible, I disagreed even more with him than with the RIAA.

Stratton responded on January 26, 2006. Here's the response, in full:

> I would not characterize myself as being anti-copyright. As a both an artist and a copyright lawyer, I'm pretty fond of copyright and of working in the field. What I advocate is reforming the law. It is not difficult for me to be in favor of reasonable copyright and to be dismissive

of authors, because I view copyright as a wholly utilitarian body of law, intended to best serve the public.

> This explains why I do not think that authors deserve anything. I think that copyright should be offered to authors as a means of causing them to create and publish works. That is, I am interested in their works, and am willing to exploit them through the means of copyright. That they do not deserve anything does not preclude a quid pro quo. However, while having any particular work created benefits the public, granting a copyright on that work harms the public. Operating on utilitarian principles, it is important to get the most benefit for the least detriment. I believe copyright law can be reformed to more closely approach this optimal point.
>
> Given the right sort of copyright law—one in which protections are likely quite reduced from where they are now both in terms of length and breadth—I would of course staunchly support the system. I would also keep a close eye on it, so that it only improves and is not abused. And I would keep an open mind, in case a radical new idea came along that better achieved the goals of a utilitarian copyright system.\
>
> Being against copyright only makes sense to me where the public would be best served without it. That is, where the benefit of having works created due to the incentive of copyright is always less than the detriment of even the slightest copyright protection. I see such a scenario as very unrealistic at best and thus I am not against copyright.
>
> I imagine that most people against copyright hold that opinion because they're so disgusted with the entire system that they cannot see how it could be changed to provide them a net benefit, or because they have not considered how a small harm now could yield a great benefit later. Fix the system sufficiently, and I think that you'd see less agitation for abolishing copyright.
>
> Anyway, I hope this has revealed what my position actually is. I hope also that I can move up in your estimation from below the RIAA. I'm happy to discuss it with you further, if you like.

To understand the letter, I had to go back to the post (more than a year ago!) and the 63 comments attached to it. Rereading the flow, I see Stratton has a point: While I didn't exactly quote him out of context, I failed to provide enough of the flow of discussion. Stratton was responding to Rob Rickner's statement that creators DESERVE and have EARNED a right to control because of their hard work (Rickner's emphasis). I agree that "hard work" is not, in and of itself, justification for the kind of copyright protection currently in place—indeed, if I do the hard work of collecting phone numbers and putting them in alphabetic order by the subscriber's last name, I neither deserve nor receive copyright protection at all.

Stratton's probably more utilitarian than I am, but I was wrong to lump him with the anti-copyright cadre or disagree with him more than RIAA. My apologies.

Jessamyn West on Creative Commons "NC"

West posted "Copyright, licensing, the government and you" at *librarian.net* on February 1, noting "What NC means to me," generally agreeing with my philosophy, and mentioning three times when the "NC" designation on the blog has come into play:

➢ The New York *Times Magazine* reprinted a text version of her "Five technically legal signs for your library," with changes and incorrect credit. She wrote a pointed email (noting the site license); the magazine published a "heavily edited response" in the next issue.

➢ A Wikipedia editor wanted to use one of those signs to illustrate the article on West; she agreed, but needed to remove the BY-NC license, because Wikipedia operates under Free Content rules. That's not a problem; nothing in a Creative Commons license prevents the holder from granting *further* rights.

➢ TechSoup wanted to reprint a West piece from WebJunction and asked for a CC license—and West's holding of a CC license "made it a little easier to have the content presented the way I wanted it to be presented."

There's more to the post, worth reading on its own.

Steven Backs on Library 2.0

I've received some nice informal feedback on the special issue (along with a touch of indirect testiness, to be sure). Here's one—with a bit of it redacted because I believe things have changed:

> Your work on Library 2.0 strikes me as particularly needed at this point. It seems to me that this topic will gain some legs, but I hope that it will also gain a little more depth as well. For instance, a colleague and I were just discussing the notion that libraries are losing young patrons because we are failing to remain relevant in a "post Google" world. Both of us were able to recall that there is a well-known pattern among teens; they use the library as kids because their parent's take them there, then they leave us until they have their own kids, starting the cycle over again.

> [One Library 2.0 person] seems to not be aware of that, or at least [he or she] does not acknowledge it, yet [he or she] stakes an awful lot on the claim that the traditional library is to blame for the lack of interest and that only Library 2.0 can win the kids back. Maybe [he or she's] on to something, but the body of work needs to gather a little depth and needs involvement from others

who can examine it against the larger context. You have done us a world of good by doing so and I applaud the effort. Thank you for doing the work.

While I believe the comment that begins the second paragraph has some truth to it, I'm also seeing recent statements by the person named that suggest a recognition that traditional libraries aren't necessarily failing (although there's little question that libraries can *and should* do more). As it happens, I admire the recent writing from the person in question, so I'm taking the editorial prerogative of hiding the name.

Net Media
Blogs, Google and Porn

Will my emailed announcements for this issue get clobbered by email filters? We'll see. I'm not making a connection between blogs and porn (I haven't looked at *Fleshbot*, so I can't say whether such a connection exists)—but there's *definitely* a recent connection between Google and porn.

The power law continues

Jean Véronis posted "Blogs: The last will never be the first" at his semi-bilingual *Technologies du langage* (aixtal.blogspot.com) on October 3, 2005. He notes his own Technorati ranking (4,724 based on links from 210 sites at that point): "Not bad for the old ego!...All the more so since the disproportion between languages means that Blogs in French are at somewhat of a disadvantage…" He looked at the relationship between Technorati rank and number of referring sites more closely, surveying "about one hundred blogs that go from one end of the ranking to the other." No surprise: The relationship roughly follows a power law. He states that nicely, for those comfortable with logarithms: "If we put the ranking on one axis, and the number of sites on another, and we put the whole thing in logarithmic coordinates, we get a more or less straight line." That's typical of Zipf distributions and most other power-law cases. He notes where the "straight line" fails in his survey, however: Past about rank 10,000, the number of links *drops* faster than a power law would suggest. "In a way, there are 'too many' blogs who have few incoming links."

He suspects that's due to spam blogs or "splogs," which one observer suggests may make up 60% of Blogger blogs. That might be true, although I'd expect the dropoff to come further to the right. In any case, he notes that the power law "can lead bloggers to de-

spair" since it means a tiny minority of blogs get nearly all the references, "while the immense majority of blogs are not quoted (or perhaps even read) by anyone, or certainly by very few people." I'm leery of the proposition that not being linked to means not being read, but never mind. The next two figures are startling but not at all improbable:

➤ "Only" about 777,700 blogs (as of last October) had references from two or more sites (that's still an enormous number).

➤ Roughly 93% of all blogs aren't referenced by anyone.

He believes "the inertia of the 'big guys'" makes it difficult for anyone to climb very far up the power law. One exception, which moved to 90th place after a few months, was largely because of heavy coverage of Katrina—and that exception slipped after the immediate crisis passed. Technorati now counts only links within the last six months, but since that includes blogrolls, I'm not sure it has much effect: The chances for a new blog—particularly one that's not a "problog" (professional blog) setting out specifically to gather a big audience—to break into the hot 100 or even the warm 1,000 are pretty small.

When he wrote this piece, you needed 552 referring sites to be in the top 1,000; 200 for the top 5,000; 120 for the top 10,000—and 20 for the top 100,000. A fair number of library-related sites belong in that broad category (a Technorati search in early April 2006 shows 39, but that's only blogs whose owners have "claimed" them and used Libraries as a tag, and some of the most popular library blogs aren't on that list), but few fit in the narrower categories (14 claimed-and-tagged blogs in the top 10,000, seven in the top 5,000—and none in the top 1,000).

Seth Finkelstein and Jon Garfunkel have written about these issues a lot, and I've discussed some of that writing in previous issues. Things haven't really changed—and probably won't. The power law is a broad phenomenon and the "echo chamber" nature of blogrolls and fandom make it more obvious within blogging than in a lot of other areas. I'm with Seth F. (in a February 14, 2006 post) in finding Technorati's "authority" feature unfortunate—it equates popularity with authority and uses that as a way to *further* reduce visibility for less well-known blogs. Technorati bills it as "a good way to refine your search results"; I'd call it a good way to avoid distinctive and unusual perspectives. Those who regard *USA Today* as the most authoritative newspaper should love the feature.

If anyone seriously claims that popularity is synonymous with authority, I would assume that they agree that it's authoritatively true that the U.S. was created a few thousand years ago with all existing species in place—after all, that's certainly the most *popular* view in the U.S.

A curious *New York* article made the blog rounds: "Blogs to riches" by Clive Thompson, issue not stated in the web version (www.newyorkmetro.com/news/media/15967). It's all about blogging as a way to make money. Naturally, it discusses Clay Shirky and the power law, the advantage of first movers, the inbred nature of the hottest blogs ("popularity breeds popularity"), and all that. But it seems to view blogging *entirely* in terms of business models: If you're not trying to make money from your blog, why are you writing it? The article goes so far as to state that the *Huffington Post*, a relatively young blog that began with big-money backing and a "full-time staff of four" to actually post, "represents a sort of death knell for the traditional blogger." The "new model for success" is corporate blogs—that is, blogs *created* by corporations. Here comes the new boss, same as the old boss.

Putting it charitably, this is narrow-minded horsepucky. It's like saying that zines don't exist because they're not started by major publishers and generally don't make money. If the only measure for success is making money—and maybe it is in New York—then it's true that 99.9% of blogs are failures, certainly including mine and almost all other library-related blogs. But making money is not the reason most *people* blog, and most blogs are created and run by *people*, not corporations. This goes beyond the "long tail" aspects of most traditional media, where a specialized journal with a circulation of 1,500 may be quite as successful as a national magazine with a circulation of 1.5 million; this goes to the zine market, where a few *dozen* readers may represent success.

It's probably important to say at this point that Seth Finkelstein and Jon Garfunkel are, as far as I can tell, right about what they call "gatekeepers"—within any given field, a relatively small number of bloggers commands most of the attention and, to some extent, dominates the topics under discussion. For relatively small fields, that may not be an awful situation: It's not too difficult to break into the top hundred library-related blogs (or even the top fifty). But, as Finkelstein notes, that's little solace if the fields you're interested in aren't narrow fields—if you're interested in politics or the like. There, things seem to be getting worse:

The chances of a single amateur to be heard aren't zero, but they're no better than in traditional media.

Attack of the blogs

As I was thinking about the *New York* article, I realized that it wouldn't have bothered me if it was in *Forbes* or *Business 2.0.* In those magazines, you'd expect money to be the only measure of success. I expected better of *New York*, but I've always been naïve.

Speaking of *Forbes*…an article with the title above (by Daniel Lyons) appeared at *Forbes.com* for November 14, 2005. It's a doozy, starting with this lead: "Web logs are the prized platform of an online lynch mob spouting liberty but spewing lies, libel and invective. Their potent allies in this pursuit include Google and Yahoo."

Pretty strong language and it doesn't say "A few web logs," it says "Web logs." Nice smear of an entire medium! It goes on with a supposed horror story: A blogger made nasty comments about the head of a company. I don't know the facts of the story, although the reporting is slanted. For example, Lyons immediately labels the bloggers campaign "long on invective and wobbly on facts," but *never* identifies factual errors. Instead, he goes on to defame blogs once again:

> Blogs started a few years ago as a simple way for people to keep online diaries. Suddenly they are the ultimate vehicle for brand-bashing, personal attacks, political extremism and smear campaigns.

Blogs are labeled as a "new and virulent strain of oratory." Somehow, revealing the Kryptonite bike-lock situation is, I guess, a smear and brand-bashing. A marketing officer says "Bloggers are more of a threat than people realize, and they are only going to get more toxic." A PR VP talks about the "potential for brand damage"—and a lawyer asserts that half of the attacks are "sponsored by competitors." We're told that *Groklaw* "exists primarily to bash software maker SCOGroup…producing laughably biased, pro-IBM coverage; its origins are a mystery."

Worse: Google and other "formidable allies" of the "online haters" "operate with government-sanctioned impunity." Lyons appears to believe that any blog host should be responsible for ensuring every blog post is fair and accurate—which also implies that Comcast should be held legally responsible for assuring that, for example, Fox News is fair and accurate at all times.

It's quite a story, all those innocent little corporations being smeared by those evil online haters. There are more examples: A CNN executive who used the word "targeted" in an off-the-record conference and had that word repeated in a blog. The executive "instantly and repeatedly denied the assertions." Lyons doesn't suggest that the assertions were false. It's clear that, from Lyons perspective, it was fine for the CNN executive to lie about what he said, but evil for the "blog hordes" to keep "wailing away" with an apparently true but "off the record" statement. And he quotes a *right-wing* blogger who complains about *left-wing* bloggers hounding that White House "reporter" who allegedly worked as a male prostitute.

Deep links, EFF's blog, had a charming response on October 28, 2005: "Attack of the printing press!" It takes pretty much the same wording but applies it to pre-Revolutionary War America and the role of printing presses in undermining the benevolent authority of the King. You'll find it at www.eff.org/deeplinks/archives/004105.php.

I'm not defending anonymous libelous attacks, but that's not what most of the *Forbes* story seemed to be about. Lyons does a fair job of smearing all blogs and suggesting that bloggers critical of corporations and businessmen are "online haters" and "virulent." The remedy? Make the hosts responsible for *assuring* the truth and *fairness* of posts. Heck, for that matter, shouldn't grocery stores that carry *Forbes* on their racks be required to assure that every article in the magazine is fair and true? After all, free speech can hurt corporations (and people, as if that mattered).

Seven deadly sins of blogging

I found this one at *GreatNexus webmaster blog* on November 19, 2005 (www.greatnexus.com/blog/85.html). Pinyo Bhulipongsanon calls these the "seven worst things a blogger can do": Use free blog hosting services, ignore the basic principles of good site design and usability, be the "jack of all trades" (blog on more than one topic), don't post regularly (the writer argues for at least one post a day), write badly, spam and steal, and fail to establish a personality. The post runs four pages (followed by 20 pages of comments); these are just normalized versions of the primary points.

Once again, *it depends on your purpose*—at least as far as the third and fourth points. (It's hard to argue *for* bad site design or usability, spamming, stealing, bad writing, or impersonal blogging—and the point about free host services is a tricky one.) Yes, if you want a big audience of people going directly to your blog, so you can get the big ad revenue, you have to post every day. But for many of us, with aggregators,

that's not what blogging is about. Most *amateur* blog-gers—most "real bloggers," if you will—want to find their appropriate audiences, people who will appreci-ate or be engaged by what they have to say. Also, yes, sticking with a narrow topic may make you more of an Expert on that topic, and if you're blogging to fish for speaking or writing invitations, that's a good thing. On the other hand, *Boing Boing* is still the #1 blog (as far as I know), and that's not what I'd call a focused blog. Some of us *like* to be surprised by the blogs we subscribe to; we're interested in what people have to say, and it doesn't hurt for new areas to show up.

Comments are all over the place—some agreeing, some pointing out problems with the blog (it uses a fixed layout, always amusing on very small and very large browser windows), some adding new points. "Not answering [to] comments" comes up as one new sin. One commenter specifically notes that the aim is "just where you intend to go with blogging"—and for a personal journal, hosted blogs make good sense.

Writing and authority

"Momus" contributed an excellent piece at *Wired News* on November 29, 2005: "Blogging with a wooden tongue." It's about PR blogs—"official websites" that *really* violate the last rule in "seven deadly sins." Tell-tale signs of a wooden-tongue blog: Content claimed to be written "by someone powerful who's obviously too busy to write a blog" that reads "like it's been phoned in"; the blog never raises controversial topics; the blogger is "incongruously humble and modest." Momus provides an example, apparently from the curator of an exhibition but with none of the *flavor* of what goes into mounting an exhibition.

InfoTangle's blog/article for February 20, 2006 is "Authority in the age of the amateur" (infotan-gle.blogsome.com, find from there), a six-page article (four pages plus 28 endnotes) that discusses some concerns raised by critics of blogs: They lack filters, they lack authority, bloggers are amateurs. I'm not entirely convinced by some of the answers—does ap-pearing on lots of blogrolls really constitute author-ity?—but it's a thoughtful discussion. After discussing whom we trust these days, the author offers sugges-tions for judging the worthiness or authority of a blog. She suggests that librarians use "their unique expertise to evaluate and recommend authoritative blogs" by creating OPML-based reading lists. An interesting ap-proach. I wonder which librarians I would trust to recommend "authoritative" blogs?

John Scalzi posted "Writing tips for non-writers who don't want to work at writing" at *Whatever* on February 12, 2006 (www.scalzi.com/whatever/004023.html). It's a nice casual discussion, but I be-lieve he gets some of the punctuation guidelines wrong (even wronger than my frequently poor punc-tuation). He *does* recognize that, while brief para-graphs may be good, it's easy to overdo it—how many online sites have nothing but single-sentence para-graphs, with meaning chopped up into Little. Sepa-rate. Bits? "Learn to friggin' spell" makes the point that every spelling error cuts 5 points from your "ap-parent IQ"—and that every mistake of the "there, they're, their" type—the ones spell-checkers won't get—drops your apparent IQ by 10 points. He notes how many MAs and PhDs are prone to such errors. He also suggests that you not use words you don't really know (particularly slang) and offers a number of other points, starting and ending with "speak what you write"—the idea that good writing should emu-late speech. Fair warning: While the post is only 6.5 print pages long, comments go on for *61 more pages.*

Lori Mortimer offered a response of sorts at *Blog-critics.org* on February 15, 2006: "One simple rule for improving your writing." (blogcritics.org/archives/2006/02/15/180927.php). I suppose the "one…rule" is the first of four guidelines Mortimer says Scalzi missed: Use the active voice. The others: Use simple, strong verbs; sleep on it; and get feedback from at least two people. Mortimer seems to advocate these ideas—even the last two—for blogs, where it strikes me they're improbable and possibly inappropriate. (How many of us "get feedback from at least two people" for *any* writing prior to publication for online writing or submission for print publication? All the time?)

Mortimer dissects Scalzi's punctuation guidance extensively. She's probably right in some areas—but some of her advice is more confounding than helpful. Scalzi's advice on periods: "When you're writing down a thought and you're at the end of that thought, put a period." That's way too simple—but what can you do with Mortimer's counsel? "The only way to know where to put a period is to know where a sentence ends. And the only way to know how a sentence ends is to learn the parts of speech, usage, and sentence construction." Gee, that helps.

Google and Friends

Gary Price posted "Keeping yourself out of web and other databases" at *Search engine watch* on October 3,

2005. He notes a *Wired news* article about a person who "values her privacy" and is trying to keep herself out of Google. "We've seen stories like this before." Price offers a reality check. (blog.searchenginewatch.com/blog/051003-152112)

First, it's not just Google—and based on my own experience, Google's spiders are not the most aggressive these days, although they used to be. (In January 2006, Yahoo! Slurp hit *Cites & Insights* 5,123 times; Googlebot a mere 2,340, not all that far ahead of MSN Robot at 1,528. The rest—and there are a *lot* of spiders out there—top out at 498 hits during the month. But then, over at *Walt at Random*, two different Googlebot "robots" seem to account for more than 10,000 hits in all, while Yahoo! Slurp accounts for a mere 4,116 and msnbot 3,585.) "Staying out of Google" will only keep you anonymous from people who've never heard of other search tools.

Beyond web search engines, as Price notes, there are lots of other tools to find out about people, both within the open web and more so within "deep web" databases. For a few bucks, you can get aggregated information from several services.

If you're "out there" it's not Google's fault—although it's true that Google and competitors could be more up-front about ways to keep material out of the databases. Beyond that, I think Price is just barely right: "Trying to remain completely and totally private in the United States might be possible. Very difficult, but I guess possible." *Just barely* possible, and probably not worth the effort.

For most businesses and bloggers and websites, the desire is different: To be as prominent as possible in Yahoo!, Google, MSN, and the rest. Some of them want that prominence to be selective—and one porn site, Perfect 10, seems to be having some success in suing Google over the issue. On February 22, 2006, Judge Howard Matz issued a preliminary injunction against Google's display of thumbnail images from Perfect 10 within Google Images. I've seen thoughtful discussion of the findings and issues from Seth Finkelstein at *Infothought*, Fred von Lohmann at *Deep links*, "kim" at *LawFont.com*, and—briefly—Alan Wexelblat at *Copyfight*. There's also commentary at *Sivacracy.net*, in the usual combative tone that site now seems to take in all manners Google-related.

I'm assuming here that Perfect 10 doesn't have a "no-crawl" file on its site, since Google would honor such an instruction. I'm guessing the porn site *wants* to be discovered via search engines. But Perfect 10 has

a business for cell-phone users who like their nekkid women on the very small screen, selling pictures through Fonestarz that may not be much larger than Google Images thumbnails. So Perfect 10 claims Google is interfering with an existing business use, helping to undermine fair use claims. They also claim Google's potential sharing of ad revenue with infringing sites that copy the Perfect 10 images, via AdSense, constitutes commercial gain.

What's bad here? The court's broad definition of "commercial use" could impact Google Book Search—and a lot of Google and its competitors. Fred von Lohmann includes a fascinating tidbit on the "market interference" portion of this: "As for Fonestarz, I don't think the court was adequately sensitive to indications that the arrangement was a sham concocted for this litigation (the court notes that the license was not entered into until after Perfect 10 sued Google)."

Otherwise, not as much as you might think. The court rejected the idea that Google linking to an infringing site *itself* constitutes an infringement, even if that link is the same-page display you find in Google Images. The court also rejected the idea that Google should be held responsible for infringing sites that it links to because it "created the audience" for those sites. And the court couldn't buy Perfect 10's notion that *visiting* an infringing website (and, thus, temporarily caching copies of infringing material) inherently constitutes infringement—a notion that would devastate the web as a whole.

The injunction is preliminary, subject to appeal—or to Google making an arrangement with ("paying off" is such an ugly term) Perfect 10.

A couple of other Google-related items may be worth noting. You've seen one of them, although you may not realize it. As John Battelle noted in September 2005 (battellemedia.com/archives/001889.php), Google did two things: Claimed that its index was three times bigger than any competitors—and stopped showing an index size claim on the home page. Since Yahoo! and MSN (the two primary competitors for web search) are both portals, their claimed index sizes aren't featured prominently in any case; now, you only see such claims on smaller sites such as Exalead. Yahoo! famously claimed last summer that its index was bigger than Google. The Google claim is as unprovable as the Yahoo! claim, but what's new is Google changing to "most comprehensive search engine by far" without numbers—and, as before, without any good way to test the assertion.

As Battelle points out, stopping the numbers race may make sense, since it returns the focus to relevance. My own sense, for my own searches, is that "relevance" has gotten worse for all three engines; that any engine handling more than a couple billion sites seems to include so much noise that it's increasingly hard to ferret out the signal. That's partly because I'm not just looking for the "top" result—I'm looking for the useful results within the range the engine is willing to show me. My recent experience? Sometimes MSN Search is better. Sometimes Yahoo! is better. Sometimes Google is better. I don't find any of them consistently best. But that's me.

Incidentally, Battelle got a response from Yahoo! about Google's decision to stop mentioning numbers:

> We congratulate Google on removing the index size number from its homepage and recognizing that it is a meaningless number. As we've said in the past, what matters is that consumers find what they are looking for and we invite Google users to compare their results to Yahoo! Search at http://search.yahoo.com.

To which Battelle responds, "Why on earth, then, did you announce that 20 billion number in the first place?" But he's happy that "this is the end of it."

Finally, on the off chance that any *C&I* reader doesn't already know about it, I should mention *Google's Newsletter for Librarians*, launched in December 2005. Full transparency: I accepted an invitation to write an article for a future issue; I don't yet know *which* future issue. You can sign up for the newsletter at www.googlelibrarian.com, which also links to a blog. The first issue featured Matt Cutt's "How does Google collect and rank results?" It's a little simplistic (and, of course, it doesn't reveal any of Google's proprietary ranking mechanisms), but it's a good start.

My Back Pages
50 Best New Sites

That's the message on the cover of the February 2006 *PC World*—but the story's called "New, improved Web." It's not "50 best new sites"—it's "a sampling of the most useful and interesting sites and services of what some call Web 2.0." If it's not Web2.0ish, a site's not a candidate no matter how wonderful. The story? Well, with lines like "Microsoft has hopped on the New Web train in the nick of time" you'll have to draw your own conclusions. This comment is mostly for the bait-and-switch cover line.

Cars in *PC Magazine*?

There it is, in the February 7, 2006 issue, with a "www.technoride.com" URL. Bill Howard offers "The technology cocoon," telling us about all the computer-based wonders that we *really should have* to lull us into thinking that driving doesn't require alertness. (No, he doesn't say that—but he does laud devices that "help drivers stay properly positioned even when they're not fully alert.") Here's the rundown: Active cruise control (but "it doesn't function above 90mph"—which is a problem *why* in the U.S.?), lane departure warning, backup sonar and backup cameras, and blind spot detection. Active cruise control doesn't help if the driver in front slams on his brakes (it only gets you down to 20mph) and "doesn't deal well with dense traffic"; lane departure warning doesn't work well in heavy rain or snow (when *awake* drivers could use it); and these devices won't do much if a hapless pedestrian walks out in front of your two-ton "cocoon." Next to the column: photos of a Lincoln Zephyr, Ford Fusion, and KVH TracVision A5 DirecTV car unit—with ratings for each one. *This* is the future of personal computing magazines?

Two Curiosities in One Page

Page 34 of the February 7, 2006 *PC Magazine*: two half-page "first looks" reviews. The upper one is for the Voodoo ARIA, a $4,897 Media Center PC (no real display for such a modest price, of course); the lower, for the Lenovo ThinkCentre M51, a $1,329 (with monitor) business PC. The oddity up top: one "really useful feature is the DVD Ripper…By launching DVD Decryption, a freeware DVD-ripping utility, we were able to copy store-bought DVDs to the hard drive." There's a little matter of DMCA here…as you'd think *PC Magazine* would know.

The bottom? Nothing wrong with the writeup as such—but there's a statement within it that I believe is true and find both bizarre and unfortunate. "Although the machine retains the antiquated floppy disk drive (IT departments still demand it)…" *Really*? In 2006? Business IT departments are insisting that business computers have microdiskette drives? After the obvious questions ("What on earth for?" and "Have these people heard of $25 USB drives and $0.10 CD-Rs?") comes a third one: Have these IT departments ever *tested* those microdiskette drives? The last diskette drives I had in PCs at work and at home, worked about half the time: Your chances of being able to read

a diskette on one PC that was written on another PC were typically about 50%. Given that the drives were basically throwaways, that's not surprising. So why would businesses want unreliable, bulky, low-capacity data transfer devices in 2006?

Blind Testing and Logic

Serious audiophiles must be aware of the controversy surrounding blind and double-blind testing—that is, component testing where the reviewer doesn't know what's being tested or, in the case of double-blind, neither the reviewer nor the person setting up the test knows which of two items being reviewed is active at any time.

Those assailing double-blind reviewing make good points. The rapid switching and short sessions typical of that process tend to mask small differences, particularly since they increase listener tension and fatigue. The differences that high-end reviewers call enormous, obvious, and worth spending an extra $20K or $50K for are typically differences that can only be understood as differences over time.

But some attacks on blind testing raise questionable issues, such as this one in a letter to *Stereophile* (February 2006):

> This whole discussion of blind testing is not applicable to music reproduction. All of us in this hobby know that each and every one of us has varying tastes in what we are looking for in sound. Some of us like articulate reproduction, while others like house-moving bass with a delicate sense of the mids.

That's part of a long discussion…which has *nothing* to do with blind testing. Someone who prefers one kind of sound should be able to determine how closely a component matches their preferences without knowing what the component is—unless, of course, their preference is really for "that Halcro sound" or "the obviously superior Musical Fidelity sound" or "whatever sound expensive components from my friends' companies make."

If you're convinced that a $350,000 Wavac amplifier really sounds better—by *your own definition* of "better"—than, say, a $500 Harman Kardon or Rotel receiver, you should be able to describe and identify that sound convincingly even if you never know that the amplifier you're listening to is a Wavac, not a Rotel or Pioneer or whatever. After taking as much time as you want. That's what I've always wanted to see—the "Pioneer box" test, in which each new piece of electronics is tested *for sound quality* while it's in a sealed box that originally held a mass-market receiver. Taking as long as the reviewer wants. I won't hold my breath until that happens.

…and Here I Thought These Components were Pricey

An industrious letter-writer in that same February 2006 *Stereophile* did a study to determine the truth of complaints the magazine keeps getting that it's preoccupied with unaffordable components. (The typical response from other readers is that auto enthusiasts don't complain when auto magazines review $300K Lamborghinis—to which there are two appropriate responses: First, such reviews are exceptions; second, $300K is 10 to 20 times the price of a mainstream car, where high-end components can cost 100 to 500 times as much as mainstream components.)

The reader built a spreadsheet of all components reviewed over a 16-issue period, a total of 235 components. The average price? A mere $7,667—and if you exclude one absurdly expensive amplifier, it drops to $6,204. Of course, that average includes $50 digital items; the average loudspeaker price was $9,448, and the "low average" amplifier was $8,023.

Some of us might find an average review price of $7,667 for stereo equipment a trifle high, maybe even unaffordable. Some might not. The most revealing number—one that says *Stereophile* isn't truly obsessed with extremely expensive stuff—is that 39% of the prices were no more than $2,000.

Cheapskate's Fine, But Aren't There Limits?

Steve Bass' "Hassle-free PC" column in the March 2006 *PC World* starts with the statement "I'm the world's biggest cheapskate: I hate wasting anything, even a little bit of space on a 50-cent CD." He goes on to describe utilities for planning burns of large groups of files so that they use CDs as efficiently as possible.

That's fine—but, Steve, if you're so cheap, *why are you using 50-cent CDs*? It's been a while since I've spent more than fifteen cents on a name brand data CD-R. These days, you can get DVD-Rs for a quarter or less in quantities of 50 or 100. That means fifty cents should buy roughly 9GB capacity. Wouldn't it make more sense to "waste" a few megabytes and arrange files logically?

Apple Says Intel Outperforms PowerPC

There's a headline I never expected to write except possibly on a special 4/1 blog post. Given how consis-

tently Apple's touted the PowerPC architecture as *inherently* superior to Intel CPU architecture, how Apple told us megahertz wasn't really relevant, and all that—well, gosh. According to *PC Magazine,* Apple now says Intel chips deliver better performance than PowerPC chips. Which you'd expect them to say, given that Apple's giving up on PowerPC. Apple will still charge more for Macs than others charge for comparably-equipped PCs; as Michael J. Miller puts it, "Apple's strategy still positions the Mac as a niche machine."

Or you could believe John C. Dvorak, who suggests Apple's getting ready to dump Mac OS X and adopt Windows instead. That sounds a little bizarre to me—but then, consider the source. (I won't say it's impossible; I will say it seems improbable.)

The Fastest PowerPC Mac Ever?

That same *PC Magazine* (March 7, 2006) has full-page reviews of two different Mac models. The iMac (Intel Core Duo) gets 4.5 dots and an Editors' Choice award; it's a well-equipped system (1GB SDRAM, 250GB disk, 128MB ATI X1600 graphics, dual-layer multiformat DVD burner, all built into Apple's 20" widescreen LCD display) for $1,799 (as tested), but Mac aficionados might wait until more programs are updated to work well with Intel.

The companion review also assigns 4.5 dots but no Editors' Choice. It's a barn burner: the Apple PowerMac G5 Quad, with two dual-core 2.5GHz PowerPC G5 processors (that's four CPUs in all), 4GB SDRAM, two 500GB drives (*not* striped as RAID 0 for faster access, oddly), a 512MB nVidia Quadro FX 5400 graphics card, dual-layer multiformat DVD burner, and 30" Cinema Display. If there's a downside, it's the price: $9,522. Or skip the huge display and pay a mere $7,023. (Or, as they say, real power users may want to upgrade the RAM to 16GB—but that will set you back another $11,900.) If you're not a graphics or video professional, this system is pure overkill, to be sure.

How Not to do Serialization

Alex at *brokentype* (www.brokentype.com) posted an insightful essay on March 15, 2006: "Serial killer: 6 rules for serializing novels online." It's about *Slate*'s attempt at serializing a novel (*The Unbinding*), and why it's such a disaster. "Somehow *Slate* has managed to excise all of the design techniques that actually make serial novels practical and fun to read online."

There's nothing new about serializing novels; it's been done since Dickens' time. Many important science fiction novels first appeared as serials (or as sets of novellas and novelettes eventually woven into novels). *Analog* still serializes novels.

The "six rules" are that a serial novel needs an RSS feed *just for the novel* that provides a list of chapters; you should be able to get email updates whenever a new chapter is posted; the page should be comfortable to read—it should follow good online type rules; you should be able to bookmark the latest chapter so you can continue reading; the text should be *accessible*; and it should *be* a serial novel—that is, a novel that lends itself to serialization (with something like a cliffhanger at the end of each installment).

Slate's attempt? Yes, there's RSS—for the whole site, of which *The Unbinding* is a sliver. No email updates. The type is *horrendous*—white text on a black background, bad spacing, and other things that caused me to give up partway through the first chapter. No good way to get directly back to the novel. It requires Flash. It isn't even a narrative text—it's "a compilation of 'found documents'—online diary entries, e-mails, surveillance reports, etc." As Alex says, "Why is it that editors think we want to spend our time sifting through documents?...Serial fiction is a well established form, there's no reason to change it."

URL: cites.boisestate.edu/civ6i6.pdf