Cites & Insights

Crawford at Large
Libraries • Policy • Technology • Media

Net Media
Google, Wikis and Media Hacks

Before proceeding to semi-organized chunks of current items about Google and wikis, a couple of stand-alone columns caught my eye—both "Media hack" pieces by Adam L. Penenberg (an assistant professor of journalism at NYU), both appearing at *Wired News*. Not that I agree with or accept what Penenberg says, but he's interesting and thought provoking.

His April 28, 2005 piece, "The new old journalism," includes the unfortunate assumption that [all?] "younger people will undoubtedly choose the web" over newspapers and "Ultimately, the printed word will die off." Although he's talking about print newspapers, he doesn't explicitly make that limit. "It's inevitable since it will be more cost-effective…to distribute news over the web and via cell phones and PDAs…" Interesting to have cost-effectiveness as the basis for inevitability; unfortunate that Penenberg sees no loss in moving from the broad, socializing, local-business-serving role of the print newspaper to the "tell me only what I want to hear about" role that web news plays in most lives (I believe).

The survey he's basing this on says that 19% of Americans 18 to 34 *do* read print newspapers, but universalisms are always tempting for columnists. What I found most noteworthy is his assertion that people *aren't* abandoning newspapers: They're abandoning the print medium. Oddly, he includes "magazines" in this assertion—and there's *no* indication that people are abandoning print magazines or avidly adopting digital versions.

He thinks it makes sense to keep teaching the skills of journalism: We'll still need reporters even if they're working entirely in net media. "[W]hen all is said and done, I still expect that each student will know how to craft a hard news lede on a tight deadline. Because whether we're talking today or 10 years ago, it's not the medium, it's the reporter." I agree with the conclusion, even if I disagree with much of the column. ("Lede" is newspaper jargon. What's a profession without jargon?)

The July 21 piece, "Web publishers eye your wallet," is a discussion of the "Balkanization of online media"—the idea that we *will* pay for internet content in the future, with all the good stuff locked behind subscription and pay-per-article doors. The source is Pat Kenealy of International Data Group. His analogy is TV, where it was free in 1955 "and two generations later most people pay for it." That's a tricky analogy, since the most frequently watched TV is *still* free, even if most of us pay someone so we don't have to fiddle with an antenna. Kenealy uses another truly odd analogy: "We got used to paying $1.50 or so at some ATMs—and that's to withdraw our own money." Maybe *you* got used to it, Pat, but millions of us at Washington Mutual and some other big banks don't intend to pay to withdraw our own money.

Inside This Issue

The big holdup, of course, is the ever-elusive micro-transaction software: Kenealy thinks we're all just waiting to pay say, $0.50 or $2 or whatever for content if the transaction's as easy as buying a magazine at a newsstand. Kenealy's fine with the idea that you lose most of your readers when you require registration

and, in the future, a subscription: The remaining readers will be more attractive to advertisers, who will then pay higher rates. That's right: Paying for your econtent won't avoid ads, but may make them more pervasive—even though nobody's quite figured out how to make anything other than tiny text ads as acceptable in net media as they are in print magazines.

As for weblogs? Kenealy's analogy, both overbroad and nicely dismissive for an oldline print publisher: "Every blogger is a rock band without a record contract." So they'll still be free. Kenealy should learn something about what "rock bands without record contracts" are doing these days with downloads, create-to-order CD-Rs, short-run CDs, and other ways of doing without the star-making machinery. They may not be able to quit their day jobs, but they also don't sell their souls to The Man.

Google Doings

Lorcan Dempsey offered some early thoughts about "the G5" (the combination of Google and five major libraries involved in the huge digitization project) in an April 16, 2005 blog post. "If large amounts of the G5 library collections are digitized, indexed and searchable then we have an index to books in all library collections. This initiative potentially improves access to all library collections, provided we have good ways of moving from the Google results into those collections." He also discusses "coverage moving forward" and the implications of the under-copyright portion of the project for the distinction between libraries' "bought" and "licensed" collections. I won't quote more and the third discussion may be somewhat moot; go to orweblog.oclc.org/archives/000632. html for the whole story.

Google, Google Scholar, and librarians
Bill Drew, Baby boomer librarian, posted this entry on April 19, discussing a report of an ACRL program—and later that day posted a longer response from Steven Bell, one of the people involved in the program. Here's Bell's statement, which Bill Drew used as the basis for his "I disagree" initial essay:

> If you care about helping your users get to the highest quality results, it's difficult to say that Google is a good model for searching in an academic context.

Now, before you shout "I LOVE GOOGLE!" or "It's good enough" or whatever, read that sentence carefully. Bell is not saying Google is worthless. He *is* saying that it may not be "a good model" for academic

searching when you're looking for the *highest-quality results*. Drew's response is that library databases are too difficult to use; that "quite often 'good enough is good enough'"; freshman papers don't need the "highest quality results"; and Google/Google Scholar results may be all that's needed. *But that's not what Bell said.*

Drew goes on to "imagine the world where Google Scholar is the interface to all of our databases and our online catalog as well as to web pages"—and, frankly, I can't imagine that working out all that well, quite apart from the logical stretches involved. Even then, Drew says, "Those with greater needs could use the separate databases"—such as those needing "the highest quality results"? The close of Drew's original post strikes me as truly odd: "All librarians that like and use Google, do not be afraid of standing up and saying so." But very few people—Bell certainly not among them—are saying you shouldn't like or use Google. The critics are saying that Google is not the be-all and end-all. Bell responded to Drew's oddly off-center attack with a long, thoughtful email (which Bell was willing to have posted) that's better read directly. Drew precedes Bell's response by saying he and Bell are "not that far apart after all." You'll find both posts in the April archives at babyboomerlibrarian.blogspot.com.

Speaking of Google Scholar, the California Digital Libraries released "UC libraries use of Google Scholar" on August 10, 2005. It summarizes the results of a quick survey on librarian and library staff use of Google Scholar within the University of California.

> The replies indicate a core of respondents do not use Google Scholar at all. Others use it rarely, instead strongly preferring licensed article databases purchased by the libraries for use in specific disciplines. Some are reluctant to use it because they are unsure of what it actually covers.

This isn't a statistical study. Boxed sets of bullets cite uses for Google Scholar, reasons for usefulness, "I don't use it because," uses of Google Scholar at public service desks and in teaching, and other comments. Seven pages provide all responses grouped by the survey questions, followed by a brief essay from UCLA, which has elevated Google Scholar to their home page. An interesting direct look at how some academic librarians are dealing with this resource.

Why Google is like Wal-Mart
Here's an odd one: Another Penenberg "Media hack" column dated April 21. It's a snide little piece, draw-

ing plausible and stretched parallels between the two firms. For example, "Alternative slogans: Wal-Mart: 'Always low wages.' Google: 'Maybe not evil, but after the IPO not so good either.'" Penenberg says Google "accounts for almost four out of five internet searches," which doesn't agree with any other reports I've seen, and also claims Google pays less than other Silicon Valley companies. (I'm reporting what Penenberg says; I'm not convinced of any of this.) The most troubling parallel, to be sure: Just as Wal-Mart insists on censored versions of some CDs and DVDs, Google blocks some sites in some countries (of necessity) and bars AdSense affiliates from criticizing Google. (Does Google *really* pay systems administators $35K? In Mountain View? I find that a little hard to believe except as an entry salary for an entry-level position—but I have zero inside information.)

What's next for Google

Speaking of "four out of five internet searches," here's an article that flatly disagrees with that number, by Charles H. Ferguson, posted in January 2005 at TechnologyReview.com (www.techreview.com/articles/05/01/issue/ferguson0105.asp?p=0). This long article (eight pages of *very* small type) discusses likely competition between Microsoft and Google. A pie chart on the second page says Google's own sites perform 38 percent of web searches, while other sites that license its technology (some of which are moving to other technologies) account for another 10 to 15 percent. Maybe 48 to 53% equals "almost four out of five" to Penenberg, but not to me. The 38% figure agrees with other metrics I've seen.

The article's interesting and challenging, reminding us of the days when Netscape's Jim Barksdale assured us Microsoft could *never* catch up with Netscape in the browser market. It's a detailed article, arguing that Google needs to establish itself as a *platform* (by promoting APIs)—and that it should avoid going after MS in the browser (and OS) arena. There seems to be an implication that either MS or Google will "win," as opposed to expanding the current Google/Yahoo! duopoly to a broader, more competitive three- or even four-way search market. Still, worth reading.

Google Print and books under copyright

Then there's the copyright flap over the library portion of Google Print. The publishers attacked and Google retreated, at least temporarily. An oddly-titled August 11, 2005 post on the Google blog, "Making

books easier to find," notes the "two new features" for publishers—one to "give us a list of the books that, if we scan them at a library, you'd like to have added immediately to your account" (which gets publishers ad revenue and directed buyers) and one to allow publishers to "tell us which books they'd prefer that we not scan if we find them in a library. To allow plenty of time to review these new options, we won't scan any in-copyright books from now until this November." That second option is the partial retreat. Not too surprisingly, publishers assailed it because it's the wrong way around: Copyright holders don't have to provide would-be infringers with lists of "things we'd like you to not infringe." This, of course, assumes that scanning entire books into a database constitutes *copying* even if those books aren't made available except in snippets—and, for a commercial entity, even that nominal level of copying may infringe copyright. Google's lawyers apparently didn't believe that was true, at least initially. Maybe they've talked to other lawyers. (One Harvard law professor believes Google would win a court fight over fair use based on the "social worth" of their scanning.)

A fight erupted at Copyfight (see the August 2005 archives at www.corante.com/copyfight). Aaron Swartz thought Google had every right to keep on scanning. Siva Vaidhyanathan disagreed based on Google's commercial status and current law.

> If copyright is to mean anything at all, then corporations may not copy entire works that they have never purchased without permission for commercial gain. I can't imagine what sort of argument—short of copyright nihilism—would justify such a radical change in copyright law.

Vaidhyanathan is no copyright maximalist. He goes on to claim that the University of Michigan, for example, *could* do such copying for its own patrons. "I wish more libraries would push their rights under copyright." As I read the library exceptions to copyright they're quite limited, but there's no question that nonprofit libraries have more leeway than corporations do. As the multipart discussion went on, we had an astonishing suggestion that Google really was sort of a library, or at least close enough ("good enough"?) for jazz—to which Vaidhyanathan had some ripe responses that suggest he knows something about what libraries and librarians are and do.

Among the many voices in this ongoing discussion, I found four particularly interesting.

- The always-thoughtful Seth Finkelstein posted "Google Print: Copyright vs. innovation vs. commercial value" at *Infothought* on August 12, 2005, noting that Google surely isn't mounting the expensive digitization effort just because it's cool but because they anticipate commercial gains. The argument brings up one of the *intrinsic* conflicts in copyright law: protection almost automatically narrows some forms of innovation. Letting Google digitize all the in-copyright books and display only search results is an innovation, but "clearly *very* dubious under copyright." Since Google stands to gain, it is a balance issue: "The technology company can't be right every time, almost by definition."

- Paul Miller posted "Google Print on hold" that same day at *Common Information Environment*. Maybe British copyright law's different, but his take is simple: "I was sad to see that Google has bowed to the whinging of publishers… I had been impressed by the breadth of their vision…and saw plenty of ways in which access to in-copyright material could have been managed to the benefit of all (including the publishers). We give in to the whinging of those with no vision all too often." Well, yes, Google Print *could* be managed to the benefit of all—but as long as published material isn't immediately and automatically part of the commons, Google doesn't get to decide that on its own. Otherwise, copyright effectively ceases to exist; I don't considers that a desirable outcome.

- Tim O'Reilly (the publisher) "defend[s] Google's approach, arguing that this is another case where old line publishers are being dragged kicking and screaming towards a future that is actually going to be good for them." Sure, Tim, but again: You don't get to tell other publishers that they *must* submit to being "dragged kicking and screaming." You can try to persuade and Google will try to do that—but their original position (that scanning to create the search index is fair use) was probably wrong, and that leaves the choice with the publishers. (Copyright spills over into so many other areas. As a balanced-copyright advocate, I'm as frustrated with those who say "Trust us, it'll be good for you" as with those who insist on 100% control over uses they never had control over before.)

- Jenn Riley of *Inquiring Librarian* discussed Google Print and related issues on August 28 and 29 (inquiringlibrarian.blogspot.com). The August 29 post is a non-lawyer's attempt to judge Google Print against the four factors of fair use as stated in section 107 of the copyright act. It's a good analysis that concludes that the fair use claim is "far from a slam dunk in either direction." The August 28 post is even more interesting: Riley wonders whether cached web pages could also be considered copyright violations and whether *indexing and abstracting*, and for that matter cataloging, could be considered infringement? I would argue that the latter questions are simple: Preparing a *description* of a copyright item is an act of intellectual creation that results in a new (copyrightable) work; it is not a derivative work. (Otherwise, every book and movie review could be considered infringement.) As to the caching question—one answer is the one Google's trying to use with Google Print: "Tell us not to, and we won't." That is, if a web search engine caches pages that have no-spider specifications or retain those caches after a site owner objects, they could be in trouble—and they don't do either one of those. Whether you can apply opt-out logic to printed books: That's another issue.

Is Yahoo! bigger than Google?

You must have heard the claim by now: The Yahoo! index now provides access to over 20 billion items. The claim was apparently first made August 8 on the Yahoo! search blog by Tim Mayer. Two days later, John Battelle reported that Google "refuted" this claim saying, "[Their] scientists are not seeing the increase claimed in the Yahoo! index." Researchers who work at the National Computer for Supercomputing Applications decided to study the situation—and released the results on August 16, only six working days later. ("A comparison of the size of the Yahoo! and Google indices," vburton.ncsa.uiuc.edu/indexsize.html)

How did they check it out? They assume there's no filtering going on and that if Yahoo's claim is true, "a series of random searches to both search engines should return more than twice as many results from

Yahoo! than Google." Ah, but they're not willing to take raw numbers, and they know both search engines refuse to return more than 1,000 results. "Any search result found to have more than 1,000 returned results on either search engine was disregarded from our sample." So how did they get lots of queries returning relatively small results? By taking a list of English words and randomly selecting two words at a time—in all, 10,012 searches.

Note that they weren't actually examining the results, which makes me wonder *why* result counts of more than 1,000 were unacceptable. The results were striking: "On average Yahoo! only returns 37.4% of the results that Google does and, in many cases, returns significantly less." In fact, these word combinations were *so* obscure that Google only returned an average of 38 results ("excluding duplicate results," by which I assume they mean "similar to these" results), where Yahoo! returned only 14. How many searches have you done on either engine in the last six months that returned results that small?

They also assert that the *actual* number of results returned was about half the estimate on Google, only one-fifth the estimate on Yahoo! Their conclusion: a user can expect, on average, to receive 166.9% more results using the Google search engine than the Yahoo! search engine… "It is the opinion of this study that Yahoo!'s claim to have a web index of over twice as many documents as Google's index is suspicious." (Who knew that studies could have opinions?)

The speedy study, syntax and all, flew around the blogosphere—**Google's still #1**! Seth Finkelstein took a careful look at how the study was done and posted his comments at *Infothought* on August 16, using the study's name as the entry name. His conclusion: "The methodology is severely flawed, with a sampling-error bias… By sampling random words, they biased the samples to files of *large words lists*. And this effect applies, to a greater or lesser extent, to *every sample*." He offers examples—and, later, realizes that the methodology returns not only large word lists but also "gibberish spam pages." In the results he looked at—cases where Yahoo! returned *no* results and Google returned significant numbers—"Every page is either a gibberish spam page or a wordlist." Unfortunately, and as usual, Finkelstein's cogent criticism of the study received nowhere near the publicity of the study itself.

Matthew Cheney and Mike Perry replaced the earlier study noted above with "A comparison of the size of the Yahoo! and Google indices" (same URL) a few days later. This time, the searches involved two random words and a *third* word preceded by "-" (that is, the first two words and NOT the third word), which the researchers claim has the effect of "excluding dictionaries and wordlists." They also threw out any query yielding fewer than 26 "actual results" on both engines.

There's a *huge* difference from the original study: Now Yahoo! "only returns 65% of the results that Google does" as opposed to 37.4%—but the researchers continue to characterize Google results as "overwhelmingly larger," an adjective that does little to convince me that the researchers have no prior agendas going into this hurried and hurriedly-redone, project. (One of the tables is clearly mislabeled, which *really* makes me wonder about the rush to publish: As stated, it shows Yahoo! returning *more* than Google—73% more, to be precise.) The conclusions are pretty much identical to the first version. Seth Finkelstein points out that their method of excluding wordlists doesn't really work very well and why that is so.

Fact is, the study could *not* conclusively prove that Yahoo! is lying (which is certainly the implication). No study could, short of actual access to both companies' server farms. There's no reason to assume that Google and Yahoo! index documents identically (e.g., how deeply they index very long documents) and every reason to assume that they do not. There's no reason to assume that they define "document" identically. There's no reason to assume that the algorithms for blocking spam pages, eliminating near-duplicates, and otherwise making results semi-manageable are identical—and every reason to assume they're not.

It would be interesting to see strong anecdotal studies using *real* search terms—understanding that even 50,000 such searches would still be anecdotal. Then again, if you can't get beyond the first 1,000 on either engine, does it matter all that much which one is larger? What might matter is which engine returns a higher percentage of highly relevant results within the pages that a typical user would scan.

Librarian to Google: Stop being evil

Here's an odd one: An August 16 post at *Nexgen librarian* by Fritz "Ian" Herrick. He believes Google is "threatening the public library" and calls that evil.

> If you needed a list of dry-cleaners in Syracuse, you used to call the library. If you needed the zip code of

an address in Tallahassee, you used to call the library. If you needed to know the capital of Mozambique, you used to call the library. Now, everybody uses Google.

Have you ever called your library for a list of local businesses—or do you use the yellow pages? Herrick thinks taxpayers will say, "Everything's in Google. Why are we paying for a library?" and be happy enough when the city cuts the library budget. I wonder how many taxpayers think ready reference is the primary benefit they receive from public libraries? Last time I looked at this situation, healthy public libraries averaged about 12 circulations per person in their service area—and considerably less than two reference transactions.

Herrick's list of what would be missed if library budgets get cut is reasonable, although he ignores one *huge* thing Google doesn't do: Circulate books, DVDs, and other materials. For free! Nearly every survey shows that the public wants *books* in their libraries. Google won't change that.

Simon at *VALISblog* responded two days later, with "Librarians to Google: stop being evil (our buggy whip sales are down)." His response (in part):

> If Google is good at answering people's factual reference questions, then let it continue to do that. Criticizing Google from the assumption that we have a divine right to continue to perform this role is arrogant.
>
> Either we need to do what we do better, or we need to stop doing it, and let Google do it. And then refocus what we mean by 'library'…the library as place…the library as entertainment source (books on paper are still better and easier to read than books on screen); the library as source of serious scholarly information… We can do things that Google will never be able to—so let's use it as a resource and an ally, and concentrate on marketing our strengths.

Wikipedia and other Wiki Watching

If you've followed some of the discussions regarding Wikipedia, you may already know about the two-part *Early history of Nupedia and Wikipedia*, written by Larry Sanger and posted on slashdot April 18 and 19, 2005. (The essay will also appear or has appeared this summer in *Open Sources 2.0*, an O'Reilly publication.) Yes, this is the same Larry Sanger who posted "Why Wikipedia must jettison its anti-elitism" at Kuro5hin, discussed in *Cites & Insights* 5:3 (February 2005).

This is a long essay, particularly by slashdot standards: Part 1 runs 26 pages (admittedly fairly narrow pages), with another 27 pages in Part 2. By April 20, when I printed off the posts and first-level comments, they already added 18 and 12 pages respectively.

Sanger is *not* anti-Wikipedia: "Wikipedia as it stands is a fantastic project…" He considers himself one of its strongest supporters, is partly responsible for founding it, "and I still love it and want only the best for it." He'd like to see it *better*, though, and that seems to disturb lots of readers. His memoir starts with Nupedia, an earlier and very different project:

> Nupedia was to be a highly reliable, peer-reviewed resource that fully appreciated and employed the efforts of subject area experts, as well as the general public. When the more free-wheeling Wikipedia took off, Nupedia was left to wither…

He believes that was unnecessary, and that a redesigned Nupedia could have worked *together* with Wikipedia to "be not only the world's largest but also the world's *most reliable* encyclopedia." He offers a brief history of that earlier project (and makes it clear that both ideas came from Jimmy Wales).

If you care about Wikipedia, it makes sense to read this memoir, since you've doubtless read some of the ecstatic writeups of Wales' genius. Sanger does not try to detract from Wales; he does offer additional perspectives.

Meredith Farkas at *Information wants to be free* set up a wiki for the ALA Annual Conference in Chicago. A post on July 5, 2005 offers observations about that wiki and what it means for future conference wikis. For example:

> 1. A wiki *must* have a specific purpose.
>
> 2. You can't just offer a wiki to the public as a blank slate and expect people t add to it…
>
> 3. It's good to add some content to the wiki before making it public…
>
> 4. You need to make it *very* clear that people can add whatever they want to the wiki or they'll ask you to do it instead of doing it themselves…
>
> 5. If your name is on the wiki, some people will email you assuming that you wrote everything on it…
>
> 6. Yes, spam is a problem, but a manageable one if you have enough loyal users…
>
> 7. It is amazing to watch what a wiki has become…

The ALA Wiki did succeed. It's still available (mere-dith.wolfwater.com/wiki/index.php?title=Main_Page) and includes what must be the most impressive set of conference reports I've ever seen—103 in all, many (most?) consisting of links to reports in blogs and elsewhere. The wiki still provides an enormous gath-ering and organizing service. As Farkas says, "It's great to have a single place to read all of the reports people have written about the conference."

Based on that success, Farkas has established an-other wiki, "Library success: a best practices wiki"—"a one-stop shop for inspiration." It has its own domain: www.libsuccess.org. Take a look. If you think some-thing's lacking or you disagree with something—well, it's a wiki. You can contribute. (I may not be a wiki contributor at this point; that doesn't mean I regard them as bad or useless. Quite the contrary.)

Both Good, Neither Good for Everything

Here, then, three pieces discussing *both* Google and Wikipedia. Stephen Manes' August 15, 2005 "Digital tools" column at *Forbes.com* is "Google isn't every-thing." Here's the first paragraph (after a tease that's pro-library, but apparently only for virtual services):

> In the age of Google, when we wonder about stuff we want instant answers. I happened to wonder about the first recorded use of the term "personal com-puter," so I Googled around and ended up at Wikipedia, the hit-or-miss user-developed encyclo-pedia, whose "personal computer" entry declared au-thoritatively that "The earliest known use of the term was in *New Scientist* magazine in 1964, in a series of articles called 'The World in 1984.'"

Manes goes on to say that he still doesn't know the answer. But he knows Wikipedia got it wrong, thanks to "an even older purveyor of information: my public library"—where he found a November 2, 1926 *New York Times* article (in an online database) quoting John W. Mauchly saying "There is no reason to suppose the average boy or girl cannot be master of a personal computer." Manes goes on to discuss all the stuff you can get for free online from your library, stuff that would cost you elsewhere. Good column; too bad Manes limits his praise for libraries to online offerings.

Laura at *Lis.dom* (lisdom.blogspot.com) posted "what for and for what," noting the need to ask "for what?" when discussing whether tools are good.

> The answer to "Is Wikipedia a good source of infor-mation?" is not "Yes" or "No"—it's "A good source of information *for what?*"

That's a sensible distinction. As Laura notes (again, I don't believe I've met her, but she signs her posts with one name), Wikipedia's probably a *great* place to find out about podcasting, but might not be the ideal source for an "analysis of gender roles in *A Winter's Tale*." There's more here and it's good: Like it or not, every "objective" source is objective with a viewpoint. Google Print and Google itself are good for some things, not for other things. "There's no such thing as a 'good source of information' or a 'good technol-ogy'—there are only sources of information and tech-nologies that are *good for* certain things." This is a fairly long post (four print pages plus comments), worth reading in the original: It was posted August 3. Jane at *A wandering eyre* (wanderingeyre.blogspot.com) wrote a followup post on August 4 pointing to the *LIS.dom* post and expanding on it a bit, and I would never disagree with this sentence: "We should learn to not only harness the technology around us, but learn to examine it critically."

The Censorware Chronicles

It's been more than a year since the last censorware roundup—mostly because not a lot has happened since the first round of post-CIPA articles and discus-sions. In some ways, that's unfortunate: Censorware *still* doesn't work—but librarians are living with it and patrons presumably just go somewhere else for the sites blocked by the software.

➤ On July 17, 2004, Mary Minow posted an es-say at *LibraryLaw Blog* urging public libraries *not* to block violent sites. "There are no public libraries filtering violent websites that have been sued that I know of. Don't be the first." Apparently librarians have asked Mary whether CIPA means they *must* block violent websites—but, you know, unlike evil sex, vio-lence doesn't seem to be harmful to children. "I believe a policy or practice blocking violent sites in a public library is likely to attract a lawsuit by a civil liberties group, and I think the library would lose." I agree with Mary's take on this: "Do I want kids to see gruesome violence? No. Do I want the state defining

and enforcing what kids can and cannot see? Again, no."

➢ An August 23, 2004 Pittsburgh *Tribune-Review* story, when CIPA-required filters were still newsworthy, pointed out that censorware *was* disrupting "hundreds of routine searches" in public libraries. It starts with a search for GasBuddy.com (a gas-price tracking site) resulting in a "sex site" block from censorware. A tween patron searched for *The Westing Game*—that search was blocked for the same reason. Mt. Lebanon's library director said patrons and librarians both found the situation frustrating, with access blocked to "at least 25 commonly used Web sites" dealing with such obscenity as pensions, arts and crafts, and Villanova University. The article also quoted the "family values" people: Andrea Lafferty of the Traditional Values Coalition said, "The American Library Association is quite sinister."

➢ A similar story appeared in the *Daily Herald* (for suburban Chicago) on September 13, 2004, headlined "Technically speaking, Web filters working." The story notes library attempts to find appropriate compromises—and Indian Trails director Trejo Meehan noted one problem with censorware: "If you filter, you're assuring everybody that you can't find anything inappropriate on our computers. There's no filter software out there that can do what it really promises to do." A Websense employee admitted that "image filtering" software—at least theirs—*only* blocks based on URLs and text, not on the content of the artwork. After all, attempting to classify images might block Boticelli paintings "because [they show] so much skin." "It's extremely difficult to classify a particular image just by looking at it… Whereas, if you look at a Web site, it has lots of images and words. It's fairly easy to classify." Easy, that is, when you're not held accountable for inaccurate results.

➢ When ALA's Office for Intellectual Freedom took a quick survey of 71 libraries in the late summer of 2004, they found "most are willing to turn off the filter quickly when asked by adults," according to Judith Krug as quoted in an *LJ* item. There are exceptions: One patron called saying that the librarian told him, "You have to tell me why you want me to disable the filter" and said the request would go to a once-a-month committee. Krug confirmed this: The library said, "That's our policy." It's also a great basis for an as-applied challenge to CIPA—except that the suit would be against the library, not CIPA, since other libraries *are* unblocking instantly.

➢ Sometimes librarians *can't* follow CIPA's guidance without violating local ordinances. An October 1, 2004 story says that Phoenix' city council voted unanimously "not to allow adults to turn off the library's Internet filter." ACLU of Arizona said it was ready to take it to court; I haven't heard anything since.

➢ This March, Utah's government got into the porn-list business. The state legislature passed a bill "aimed at protecting children from Internet pornography" (but drawn so broadly that it's almost certainly open to challenge). Part of the law is an "Adult content registry" and—as Seth Finkelstein pointed out at *Infothought*—the list is to be made "available for public dissemination" in an age-restricted but non-encrypted manner. "Your tax dollars at…work?"

➢ Some of us hoped a CIPA as-applied challenge would arise in Rhode Island after ACLU of Rhode Island released "Reader's block," an 18-page report on their study of the state's public libraries. The report notes that some libraries are blocking categories such as "gambling" and "illegal," that the *minimum* statewide blocking level considerably exceeds CIPA requirements, that some libraries don't really inform patrons that they can request unblocking (although others do an excellent job in this regard), and that the biggest public library in the state routinely denies adults access to blocked material. While the report caused a flurry of blog posts and other responses and apparently caused some Rhode Island libraries to make unblocking easier, ACLU did not mount an as-applied challenge.

➢ Then there's the recent *Nitke v. Ashcroft* decision. This one's convoluted enough that you're better off going straight to *Infothought*, since Seth Finkelstein was an expert witness and has links to various resources. (Look for

July archives.) The case has to do with CDA (Communications Decency Act), yet another "decency" act. The court agreed that the act was chilling protected speech—but ruled that the plaintiff hadn't "met the burden of proof" because she couldn't prove *how much* speech was being chilled. This is an astonishing new rule: It's OK for legislators to take *some* of your Constitutional rights, if you can't prove they've taken *too much*. The decision also involves "community standards" and jurisdiction-shopping: The government can choose any venue to prosecute Internet-based speech, looking for a "community" that will find its "standards" violated. The decision will be appealed. Meanwhile, the blog entry or internet essay you write in Berkeley, Austin or Madison may be judged by the community standards of the most right-wing community in the nation.

That's it. A hodgepodge of troublesome items, mostly chipping away at freedom of speech and expression bit by tiny bit.

Perspective

Future Libraries: Dreams, Madness & Reality, 10 Years Later

It hasn't *really* been 10 years. *Future Libraries: Dreams, Madness & Reality* (henceforth *FLDMR*, since there's an entirely different *Future Libraries*) was published in January 1995. The book was written and edited during 1994, with some work in 1993. I sent the final camera-ready pages to ALA Editions in the fall of 1994. In terms of content, it's closer to 11 years.

The book's still in print. Although there was brief discussion a couple of years ago of doing a second edition or follow-up, that's not likely to happen unless one author takes it on and gets permission from the other author.

To be honest, I haven't really looked at *FLDMR* since 1998, when I was writing *Being Analog*. I rashly promised to take a hard, honest look at *FLDMR* a decade later. I've been putting it off for most of the year. Finally, I started going through the book looking for high points and flat-out errors, as well as places where

I'd say it a lot differently now. I did it the easy way: one chapter a day. (Yes, the book's an easy read. Yes, the book's short enough to read in one long sitting. But it's summer, I'm lazy, and old episodes of *Moonlighting*, *Remington Steele*, and *Gilmore Girls*—and, for that matter, *The Greatest American Hero*—beckon in the evening.)

Here's what I found, trying my damnedest to be a critical reader rather than a coauthor.

Chapter by Chapter

Going through the first chapter, I found myself checking off pages where I'd say pretty much the same thing we said back then. One exception might be this paragraph (p. 10):

> Looked at objectively, the relative roles of electronic communication and non-electronic communication (print, sound recordings, film/video, etc.) become clear. Electronic methods are best for "housekeeping" and for giving access to data and small, discrete packets of textual, numeric, and visual information (such as those found in many reference works). Each of the other media has areas in which it is the best. In particular, print-on-paper is and will be the preeminent medium for the communication of cumulative knowledge.

I'll stand by the last sentence, but the digital/analog split has become fuzzy over time. Specifically, the digital realm as just-in-time *distribution* method for medium-length narrative in the form of journal and magazine articles has proven far more important to libraries than we could have guessed in 1994. Some would claim that most readers *read* those journal articles on screen. I believe many students skim through articles in electronic form to find chunks to cut and paste, and that they may glean reasonably good understanding of the sense of the articles. For all I know, maybe KTD really *are* different and do gain full comprehension from the screen while multitasking up a storm, although I'm still not convinced.

2. *The Life of Print*
Here's a mixed case. Some magazines have fallen on hard times, although the magazine industry is still healthy. Of examples on page 17, *PC Magazine* still circulates more than a million copies but of a much less thick magazine and *PC/Computing* has vanished.

The discussion of "Appropriate technology" (18-19) is dated—but less so than you'd expect after a decade. There are readable electronic displays that don't use transmitted light, but so far they're not all

that wonderful. Resolution has improved in some cases: Some notebooks and PDAs have displays with 150 to 200dpi resolution—but that's an improvement from 100dpi or so a decade ago, and you only see ultra-high resolution on *very* small displays. At this rate, who knows when (or if) we'll get to 300-600dpi displays at least as large as a book page? Meanwhile, almost all home laser printers now offer 600dpi printing or better, as do most inkjet printers.

An anecdote on p. 24 has been a source of misunderstanding, as my coauthor used it as an argument against the worth of Google Print. I still maintain that good nonfiction books should inform well enough that you're better able to understand p. 154 after you've read pages 1-153—but that does not negate the worth of seeing a snippet of p. 154 as a clue to the usefulness of a book. If "random accumulations of opinion, disconnected data, unverified assertions, and contextless statements" was my writing, I apologize: While that's a fair description of millions of web pages, there are also millions that provide clear worth.

My discussion of the startup costs of book printing and CD-ROM production was right then but is no longer quite true. With PoD technology, startup costs for a short-run book can be nearly zero. Similarly, short-run CD-ROM and audio CD production is now CD-R production, and there are no real startup costs other than getting the content and organization right.

We discussed *PC Magazine* and its CD-ROM version and some other examples of magazine economics. As already noted, *PC Magazine* issues don't "average 500 pages" these days (150 pages is closer to the mark)—and the CD-ROM version has disappeared, I believe. *PC Week* seems to have disappeared into the digital mists.

On later pages, discussing the *Daily Me* (a concept that still disturbs me), we overestimated electronic distribution costs. For one reason or another, much of this stuff *has* stayed "free" (well, $20 to $50 a month…) although various business pundits still see most professional content disappearing behind subscription or pay-per-article walls.

3. The Madness of Technolust
What did we get wrong here? In a footnote on p.37, we thought that Apple's Newton might "succeed quite nicely" in some niche markets even after failing as a mass-market technology. Oops.

We were correct in believing that CD-I and CD-V wouldn't explode onto the marketplace. The "2.2

computers in every household by 1999" projection is still an absurd claim in 2005. "Digital convergence" hasn't happened—but we did miss the "explosion" of DVD (faster than most new media, but still a slow explosion).

Looking at dollar figures on p. 40, it's worth noting that LCDs are *finally* cheap enough to drive CRTs off the market (slowly, ever so slowly): The 19" Sony LCD at which I'm writing this cost me $500, and a cheaper unit would have been $350, where in 1995 you'd pay $6,000 for an 11" LCD color display. I'll stand by the last sentence in that comparison: "Only someone with a terminal case of technolust would trade in a 17-inch CRT, or even a 15-inch CRT, for a much more expensive 11-inch LCD screen." The gap between CRT and thin-screen devices finally shrank, to the point that LCDs are better for most computer purposes other than gaming and high-end color work. (But then, we said that on p. 47: "Is an LCD screen better than a CRT? All else being equal it may be…"—and now "all else" is pretty much equal.)

We were too optimistic about LaserVision's survival as a long-term consumer success; the discs were too big and too expensive to produce, making it easy for DVD to kill off the medium. We were right about that "dead duck" hard disk storage and its continued price advantage over RAM. Back then, the big news was that engineers got RAM down under $100 a megabyte—but hard disks were down to $0.60 to $1.00 a megabyte. Look at what 11 years has wrought: A 512MB flash RAM USB drive goes for $50 (about $100 per *gigabyte*, one-one-thousandth of the 1994 price)—but hard disks go for $0.60 to $1 per *gigabyte*, also one-one-thousandth of the 1994 price.

A couple of reviewers suggested that we were engaged in polemics. That's true. This chapter may be the most polemic of them all. By and large, it stands up pretty well.

4. Electronic Publishing and Distribution
Remember Project Xanadu? It's now been two decades of "coming to town any day now." Meanwhile, Project Gutenberg's Michael Hart continues to damage a useful project through his crazed mathematics and other publicity-hound quirks. New forms of "electronic broadside" have emerged and they've moved beyond the elite, what with blogs and wikis as well as tens of millions of personal websites. E-journals continue to emerge, but they've hardly served as "savior for libraries concerned with STM fields."

The good news is that, so far, we were wrong about "metering the internet"—most of it continues to be available for a flat monthly access fee. Otherwise, this chapter is just fine. We missed the rise of journal aggregators—but we were on the money about $20/month "virtual libraries" not catching on within the average household.

5. Coping with Electronic Information

Almost this entire chapter reads well today, in my opinion. The section "Will full-text searching in very large databases ever work?" has been partly answered, but only partly. Yes, people do it. Yes, it can be made reasonably fast—but I question whether such full-text searching produces results anywhere near as good as those yielded through professional indexing and abstracting. We suggested that computers were ill suited to determine *meaning*. I'll argue that this continues to be true.

6. Deconstructing Dreams of the All-Electronic Future

Another polemic, at least in part. The claimed projections on p. 86-87 were not straw men then, and aren't now. They may be simplifications, but they represent real statements by real, and seemingly knowledgeable, writers and speakers. Some of these statements still arise from time to time, although the tide has shifted, possibly due in part to *FLDMR*.

Pages 90-92 may be wrong: Costs for conversion have come down considerably and the cost of storage has come down even further. That section wasn't wrong in 1995, but times have changed. Still, despite the promise of Google Print, we don't know what it would cost to do truly large-scale conversion to true preservation quality, or whether it's feasible to do so on a worldwide scale. In 1995, three terabytes was a *lot* of storage. Now? Six inexpensive PC disk drives will do it, most likely for less than $3,000.

Page 98: While it's still true that "one cannot buy a computer today for $47 that is as useful as the $3,000 computer was nine years ago," $1,000 is no longer "the lowest viable price for a complete computer system if a minimal level of support and construction quality is to be achieved." That price now appears to be $400 to $500 (including display), and as little as $600 for a notebook computer—but that's partly because "minimal level of support" means something different in 2005, something closer to "it's on the web; go find it yourself."

One suggested price hasn't changed all that much: the estimate of $0.025 per page (for paper and supplies) for "individual printing." Except that nowadays the individual is more likely to use an inkjet printer—and it's tough to print text pages on an inkjet for much less than $0.04 per page including paper.

The discussion of distributed printing costs continues to be relevant (and ignored) today. Truly "free electronic distribution" continues to be at least partly a myth. Much of the polemic still stands.

7. Enemies of the Library

My sole note after reading this chapter again, ten years later: "All jes' fine."

8. The Diversity of Libraries

On the downside, there's possibly too much repetition here—both within the chapter and from previous chapters. It's no longer true that 200 million Americans lack access to the internet, but still true that "tens of millions of Americans may have no need for, or interest in, such access"—to the chagrin of those who believe you can't *live* without being online.

9. Economics of Collection and Access

Here again, we missed aggregators. That affects some of the discussion on p. 136 and pp. 147-148. Otherwise, this chapter continues to be pertinent.

10. Survival Guide to the Serials Crisis

I look at "Protecting Intellectual Property," pp. 160-163, and take some shared pride, as I do in the rest of the chapter. There was no Open Access Movement in 1995. I have no reason to believe that any of OA's founders read *FLDMR*. But I think we said important things about the serials crisis, and I think they're still relevant.

11-12. Future Libraries: Beyond the Walls; Successful Libraries Make Their Own Luck

Both of these chapters still work.

Conclusion

I was looking for problems as objectively as I could. But I'm still the coauthor. I didn't concern myself with excessive rhetoric. In 1994, there were far too many people projecting the death of print and the virtualization of the library, justifying some extreme rhetoric in response.

I'm not thrilled about the *look* of the book. For some reason, possibly the paper used, the desktop typography didn't come out as crisp and clean as I would have liked. If you compare *Being Analog*, pre-

pared with the same typefaces on the same printer using the same software and only slightly tweaked styles, you'll see the difference. (*First Have Something to Say* is even nicer looking but I didn't prepare camera-ready copy—and, as some of you may know, *FHSTS* resulted in my purchase of Berkeley Book, the type now used for the words you're reading.)

We didn't prophesy every new technological development in libraries, but that wasn't our aim. I, for one, would have done very badly in such an attempt—partly because I'm too much of an optimist. See our comment on Apple's Newton as an example.

If I had it to do over again, given 1994's situation, would I tone it down? Not really. I believe *FLDMR* changed the discussion among librarians. I'd like to think it helped move common library wisdom to recognition of complex libraries as the most likely and best future.

We did good. Sorry to disappoint those of you hoping I'd trash my own earlier work: Not this time.

PC Progress, March-September 2005

Desktop Computers

Combined reviews of desktop and notebook PCs show up here as well—such as an odd "ultimate & affordable" roundup [C25:4]. It's another "classify the user and slot a PC to suit them" article with no ratings but an "ultimate" and "affordable" choice identified in each of five categories. Hey, if you're a multimedia professional you can spend $15,597 on an Apple PowerMac G5—but you're mostly paying for two 30" Apple Cinema HD displays. Or you can spend a "mere" $3,818 for Velocity Micro's ProMagix A/V/D, a slightly more modest solution. (Both have 500GB of disk storage and several GB of RAM.)

Another *PC Magazine* mixed desktop and notebook roundup, this time of "value PCs": Desktops costing $700 or less with a display, $600 or less without one, and notebooks coming in under $1,100. Editors' Choice among desktops is the Polywell Poly 2800NF2-MX: $699 buys a 2GHz Sempron 2800+, 512MB SDRAM, 120GB hard disk, nVidia GeForce4 MX440 graphics, dual-layer multiformat DVD burner, 17" CRT, and a fair amount of expansion space. Two notebooks share Editors' Choice honors. If weight matters, Averatec offers the AV3270-EE1 for $949: 4.3lb. with a 1.6GHz Mobile Sempron 2800+, 512MB SDRAM, 60GB hard disk, VIA 53G graphics with 32MB dedicated RAM, multiformat DVD burner, 802.11g, and a 12.1" screen. If you want more power and don't care as much about weight, Dell's $999 Inspiron 6000 ($999 with Dell's E-Value Code 1111-i6000PC) uses a 1.6GHz Pentium M 730, 512MB SDRAM, 60GB hard disk, Intel GMA 900 GM graphics with 128MB RAM, multiformat DVD burner, 802.11G, and a 15.4" widescreen LCD—but it does weigh 7.0 pounds. On the other hand, it offers almost 6 hour battery life.

A *PC World* roundup of ten "space-saving PCs" [W23:6] yields three Best Buys (one per category) and six four-star ratings in all, a very strong showing. Best Buys go to the $1,611 Shuttle Computer XPC i8600b in the mini-box category, Dell's $1,403 OptiPlex SX280 as a "thin-profile" system, and—in a change from recent roundups—Gateway's Profile 5.5C ($1,885) as an all-in-one. Gateway's latest Profile design puts the PC works in the base of the LCD display instead of using a thick display casing.

PCs designed as media centers keep getting better. This mini-roundup [P24:12] covers four good units costing $1,000 to $3,500; Editors' Choice goes to Sony's $1,800 VAIO VGC-RA830G, with a Pentium 4 500J CPU (3.4GHz), 1GB SDRAM, two 160GB disks in RAID-0 configuration, a high-end graphics card with 256MB RAM, a dual-layer multiformat DVD burner (and second DVD-ROM drive), good Logitech 5.1-channel speaker systems, and—in a departure for Sony—Windows XP Media Center Edition. The keyboard and mouse are wireless. $1,800 does *not* include a display.

Digital Cameras

PC Magazine updates its roundup of "superzooms," cameras with at least 10x optical zoom lenses, in this "First looks" roundup [P24:5]. The cameras cost $360 to $1,000. Editors' Choices are two Panasonic Lumix cameras, the $500 DMC-FZ15 and $600 DMC-FZ20. The first is a 4MP camera that only needs 1.5 seconds between flash shots; the second, a 5MP unit that needs 3.5 seconds between flash shots. Both offer 12X optical zoom with a Leica 36mm to 432mm lens.

Displays

Big LCD displays keep getting more reasonable in price and performance. This roundup [P24:6] includes ten models measuring 20 to 23 inches, all with at least 1600x1200 resolution (two have 1920x1200 resolution). Editors' Choice is the Samsung SyncMaster 213T, "reasonably priced" at $900 and with outstanding performance—except that it tends to smear a lot on fast-moving images (even worse than some other 25ms. displays). Runners-up are Dell's $749 UltraSharp 2001FP and LG's $750 Flatron L2013P.

Input Devices

If you're not satisfied with the mouse and keyboard you have—and if it's not an optical mouse, you probably shouldn't be—here's a roundup of ten input devices: three mice, four keyboards, three sets[P24:9]. Editors' Choice among the mice isn't one—it's the $200 Contour Design RollerMouse PRO, a gel-filled wrist rest with a pointer control, roller bar, scroll wheel, and five buttons that sit below the keyboard (apparently centered below the standard keyboard portion). Among sets, top choice is the Microsoft Optical Desktop Elite for Bluetooth ($155), a refinement of Microsoft's wireless set that uses Bluetooth instead of an RF receiver.

Notebook Computers

This mini-roundup looks at three "Sonoma" laptops—the newest version of Intel's Centrino platform [W23:4]. Supposedly, the platform will allow for future improvements. For now, it doesn't improve speed much but also doesn't carry much of a price premium. Oddly, *PC World* chose to rate the preproduction Dell Inspiron 6000 ($2,564) but not the Compaq Nc6230. The only production unit, Gateway's M460, offers one glimpse of what a full-fledged desktop replacement costs and offers (it rated just behind the Dell, at 3.5 stars). For $2,184, you get a 2.13GHz Pentium M 770, 550MB of 533MHz DDR2 RAM, an 80GB hard drive, DVD burner, wide-screen 15.4" display with 1280x800 resolution driven by the ATI Mobility Radeon X600 with 128MB of display RAM, USB, S-Video, and FireWire ports (but no legacy serial or parallel ports), and 802.11a/b/g Wi-Fi. Battery life was only 3:17, but this is a desktop replacement.

Multimedia notebooks? I guess that makes sense these days. This roundup includes a baker's dozen "luxury notebooks."[P24:12]. The price range is $1,200 to $3,000 and the review splits units into three categories: high-end, mainstream, and "multimedia minis." Editors' Choice among mainstream units is the $1,349 HP Pavilion dv4000, beating Apple's PowerBook G4; it's reasonably lightweight (6.4lb.), inexpensive, and has a great 15.4" screen (you get 512MB RAM, 80GB disk, and a multiformat DVD burner). The high-end choice is the $3,000 Toshiba Qosmio G25-AV513, "a beast in terms of weight [9.4lb]…storage space (120GB) and connectivity." It also has a bright 17" screen, 1GB RAM, and a dual-layer multiformat DVD burner—but you get crappy battery life. (The story makes it sound a little worse than it is: "a disappointing 1 minute 59 seconds." Now *that*'s bad battery life! It's really one hour and 59 minutes, still pretty low.) Finally, Toshiba's $2,100 libretto U100 is the "mini" of choice at 2.1lb. and 8.3x5.7x1.2"—but in addition to the high price, you get a small screen (7.2") and undersize keyboard.

Optical Drives

Double-layer DVD burners seem to be taking over rapidly, with no real cost premium and almost universal multiformat capability (everything except DVD-RAM). This test of 10 burners [W23:4] shows two Best Buys, neither of which is top-rated. That's Plextor's $130 PX-716A. Of the Best Buys, Toshiba's $130 SD-R5372 has the best all-around performance (but limited bundled software) while BenQ's $90 DW1620 offers good performance at a very good price.

Three external drives add analog-to-digital video capture, making it easy to copy a videotape to a DVD—in one case, without using the PC if you choose[P24:4]. While that feature could make Sony's $300 DVDirect VRD-VC10 the most flexible of the group, there's a problem: Sony's video capture *won't* work with a PC. It's either a DVD burner or a straight transfer device. Given that, it's not surprising that HP's $200 DVD Movie Writer dc5000 gets the Editors' Choice: it's the least expensive in the roundup, offers fine recording quality, and includes a complete software bundle. All three units record dual-layer DVD+R DL discs and record DVD+R at 16X. The HP can also record DVD-R at 16X.

Portable Players and Related Devices

I guess this is a "PC category" now, for better or worse. This roundup [P24:12] includes five small-hard-drive units and five more with large hard drives.

The Apple iPod mini and Apple iPod earn Editors' Choices in the two categories—but the $200 Creative Zen Micro also gets the nod among small-drive machines (5GB). The Zen includes FM radio and voice recordings and nine different case colors with a form factor slightly shorter and thicker than the iPod mini: 3.3x2.0x0.7" as compared to 3.6x2.0x0.5".

The same *PC Magazine* [P24:12] reviews earphones for portable players, looking at sound quality as well as noise-suppression quality. There are two ways to suppress noise: active suppression vis out-of-phase sounds that match what an included microphone picks up (within limits), or in-ear phones that fit so snugly they block noise. There's an Editors' Choice in each category. The $300 Bose QuietComfort 2 phones don't perform that well as headphones (weak low midrange and awful high-frequency response) but are great for active noise suppression. They're light as big over-the-ear headphones go but they're bulky. The $130 Etymotic ER6i ear-canal phones block noise by fitting tightly and offer good frequency response, if perhaps a bit light on the bass.

PC World also reviews replacement headphones [W23:7], including half a dozen over-the-ear units and another six in-the-ear phones. Best Buy for the over-the-ear headphones is the $150 Bose TriPort for comfortable fit and good sound, although the phones are large, don't fold, and don't come with a case. The $120 Etymotic ER-6isolator gets the nod for in-the-ear units (presumably the same as *PC Magazine*'s ER6i); the only complaints are that it comes with only two pairs of tips. The Bose QuietComfort got a low rating (two of five stars) because they didn't sound very good. A sidebar discusses noise cancellation, noting that ear-canal phones like the Etymotic do a better job than active noise-suppression units.

Sound & Vision 70:6 (July/August 2005) reviews five portable multimedia players—without much in the way of objective tests, but at least with some care and detailed specifications. There's no clearcut winner, but the $500 Archos AV400 may be the most plausible of the group. It has a 320x240-pixel 3.5" screen, 20GB disk (you can get 80GB or 100GB models), weighs 9.75oz., and measures 5x3.25x0.875". Archos (and NHJ, another maker in the roundup) was perfectly happy to dub commercial movies from regular DVD players. That same issue reviews four current legal music download services (with sidebars on some others and on audio quality), concluding that Yahoo

Music Unlimited now has an edge over the iTunes Music Store, partly because of $0.79 permanent downloads (if you're a subscriber) and $60/year subscription prices for unlimited usage. Yahoo! now owns MusicMatch, my ripper/manager/burner of choice. Then there's AllofMP3: a Russian site with an English-language option that offers, well, what it says, with incredibly low prices and various downloading options—for example, $0.36 for a 320kbps MP3 download. The site's supposedly legal in Russia; whether you're infringing copyright if you use it in the U.S. is another question.

Printers

What makes a photo printer? As several roundups have demonstrated, mostly *calling* a printer a photo printer (except for specialized units that only produce photos). This roundup [W23:4] includes seven desktop printers and four "snapshot" printers—and compared their ink and paper costs to each other and, briefly, to store processing and online photo printing. With one exception—Epson's snapshot-size PictureMate, which yields a cost of $0.23 per 4x6 print at discounted supplies prices—all these units cost more for 4x6 prints than either stores (Ritz, Walgreen's, and Target averaged $0.28) or online services ($0.35 including shipping for 20-print orders). The other snapshot printers cost $0.50 to $0.81 cents per 4x6; full-size units ran $0.46 to $0.97, averaging $0.66. Where inkjet printing on photo-quality paper comes out better is when you're doing 8x10 blowups: I figure the cost at around $1 to $1.50 each, which is less than you'd typically pay for 8x10s from a photo shop. Best Buys in the review are Canon's $230 Pixma IP4000R and Epson's $199 PictureMate. Speed ranges enormously: Canon's Pixma IP8500 takes 39 seconds to do a 4x6 photo, while the PictureMate takes 136 seconds, more than three times as long.

Multifunction printers have established themselves as great ways to add a high-quality scanner and copier to your home office without spending much more than for a good inkjet printer or taking up any more space. (I only need a copier about once a month—but it sure is nice to have one *immediately* handy when I do need it, particularly one that makes nearly perfect color copies.) But inkjet printing still takes longer and costs more than laser printing, and for lots of purposes color is overkill. Now several companies offer *laser* multifunction printers. This

roundup [P24:4] includes eleven of them (ten monochrome, one color). Expect to pay $400 to $1,300, with most units in the $500 to $700 range. Expect text printing at anywhere from 12 to 21 *real* pages per minute (actual test results for a 50-page Word document, excluding the one color unit, which managed 8 ppm). Editors' Choice is the $600 Brother MFC-8840DN, with a heavy duty cycle (20,000 pages per month), built-in Ethernet and optional wireless, and very good speed and quality. It printed the 50-page document in 2 minutes and 31 seconds and spit out a 4-page PowerPoint presentation in 22 seconds flat.

PC World's multifunction laser roundup includes four monochrome and three color units [W23:7], ranging from $400 to $500 mono, $1000 or $3000 color. Best Buys are the $490 Brother MFC-8840DN for monochrome, the $1000 Canon ImageClass MF8170C for color.

Worth noting: The April 12, *PC Magazine* confers Editors' Choice status on three new devices in "First Looks" reviews: the $300 Canon Pixma MP780 Photo All-in-One as the multifunction printer of choice if you don't need networking; Xerox' $3,399 Phaser 5500/DN as a departmental laser printer (50ppm, 300,000 page per month duty cycle, excellent output), and Canon's $400 CanoScan 9950F as a flatbed scanner with first-rate slide-scanning features.

This group review [P24:11] covers seven "top of the line" photo printers, costing $200 to $550. The worst score in the roundup is four dots, a very high score. Two Canons earn Editors' Choice. The $500 i9900 Photo Printer (which can print photos up to 13x19") for its fast, high-quality photo output; it uses eight colors. The $350 Canon Pixma iP8500 Photo Printer is also fast, produces excellent photos, and uses eight colors, but only handles up to legal-size paper. Oddly enough, it scores *lower* than the two honorable mentions: Epson's $550 Stylus Photo R1800, which uses 8 colors, prints up to 13x44 inches and prints directly on optical discs, but is slower than the Canons for photos, and HP's $200 Photosmart 8150 Photo Printer, a six-color printer that maxes out with legal paper. If you plan to use one printer for all purposes, the editors say the Epson may be the best choice: it's faster for business printing.

Projectors

Lighter, brighter, cheaper: That continues to be the trend for business data projectors (which can also be used as home TV projectors if you have the right ancillary equipment). This roundup [P24:13] covers eight units over a wide price range ($800 to $2,500) and fairly wide weight range (2.8lb. to 7.1lb.). Two Editors' Choices emerge: Dell 1100MP ($800), a surprisingly impressive bargain unit (4.9lb., 480:1 measured contrast ratio, 1429 measured lumens, 800x600) and the Plus V-332 ($1,995), the lightest (2.8lb.) unit with the best measured contrast (1179:1), albeit less bright than some (1027 measured lumens, 1024x768). It's one of few data projectors that can be packed immediately after use; it also has an iris to protect the lens when not in use.

Utility Software

One problem with anti-spyware software is that reviewers don't seem to agree on the best approaches, typically resulting in widely varying results. This roundup [W23:4] covers nine programs, with Sunbelt's $20 CounterSpy 1.0 earning the Best Buy. Second is Webroot's $30 Spy Sweeper 3.2, the winner in some previous roundups. If you're not willing to pay, Lavasoft Ad-Aware and Safer Networking Spybot Search & Destroy make a good team. (I've used Spybot for some time, recently adding Spy Sweeper.) One promising sign: Microsoft's Windows AntiSpyware (acquired from Giant Software), in beta, looks to be a fine performer.

A followup [W23:8] includes three new versions and gives a four-star rating to Spy Sweeper 4—but calls CounterSpy 1.5 "a clear winner" even though it's beta software and thus not eligible for a rating.

PC Magazine reviews seven spyware applications with some cautionary notes in a typically strong introductory essay [P24:13]. Editors' Choice among this group is PC Tools' $30 Spyware Doctor 3.2, just edging out Spy Sweeper 4.

This quick roundup of security suites [C25:4] gives a surprisingly low rating to Norton Internet Security (slow antispam, disappointing censorware, and they find that Norton slows their system) and yields *two* Editors' Choices out of four suites reviewed. PC-Cillin Internet Security 2005 ($50) gets the highest score for fast scanning and first-rate tools (but its spyware detection is turned off by default); ZoneAlarm Security Suite 5.5 ($70) is a close second, and includes a unique privacy-control feature.

Another antispam roundup [P24:7] covers four fairly new tools and awards two Editors' Choices.

OnlyMyEmail Personal costs $3 per month for a special email account that filters up to three existing accounts. Spam, viruses, and the rest are stopped at the servers and never get to your computer, so downloading email is fast—and the performance was the most accurate I've ever seen, blocking 0.6% of good mail and missing 0.4% of spam in *PC Magazine*'s extensive test suite. Qurb 3.0 costs $30 and integrates with Outlook or Outlook Express; it's a brute force system that works with a whitelist—that is, all mail from senders not on the whitelist gets quarantined. Qurb attempts to verify the sender's address using Sender Policy Framework.

PC World's midlength utility roundup covers four utility suites (not security suites) and 18 standalone utilities(W23:6). This time, Symantec's $100 Norton SystemWorks 2005 Premier earns the Best Buy for its balance of features and usability, although it is resource-intensive. I don't see any Best Buy designations among the "small utilities," but a few got at least four stars out of five: Free Undelete, Acronis TrueImage (disk imaging for data recovery), Window Washer (to erase your history), PC Wizard 2005 (comprehensive system information), X-Setup Pro (system tweaks), Folder View, VisualRoute 2005 (a network trace-routing program), ClipCache (the clipboard on steroids), and Insert Toggle Key (making Windows beep when you hit Insert, so you don't accidentally go into Word overwrite mode).

Web Browsers

PC Magazine reviews mid-2005 versions of the popular Windows-based browsers[P24:10]. Editors' Choice goes to Firefox 1.0.3, for the usual reasons. Second place is the paid version of Opera 8.0, largely because Netscape 8.0 has become "cluttered and confusing."

Wireless Routers

"Unwired for speed" [P24:10] considers "fast, faster, fastest" wireless networking—802.11g, enhanced 802.11g, and "MIMO technology," the predecessor to 802.11n (coming in 1997?). The two faster technologies work—but at the expense of compatibility. Interestingly, Editors' Choice in each category came from Linksys: the $69 Wireless-G WRT54G for 80211.g, the $89 WRT54GS for enhance 802.11g, and the $199 WRT54GX for MIMO performance.

A similar roundup [W23:8] includes six units, all of them "MIMO" of one sort or another. While Link-

sys did well here also, getting one of two four-star ratings (it's hard to tell whether it's the same model), Belkin gets the Best Buy—at $238 for the combination of Pre-N Router and Pre-N Notebook Network Card (for comparison, the Linksys combo is $274).

Trends & Quick Takes
The Half-Terabyte Notebook

According to a "Pipeline" item in the August 9, 2005 *PC Magazine*, we can expect that in three to five years, thanks to perpendicular recording technology. Seagate plans to shift all its drives to the new technology and "foresees fivefold increases in capacity" during that period. In early 2006, expect a 160GB 2.5" drive. A couple years later, don't be surprised by *single* desktop 2.5TB drives: That's terabytes, although it almost certainly means 2.5 trillion characters (a lower figure by roughly 10%; technically, a terabyte should be 1,099,511,600 characters, that is, 1024^4). Hitachi's aiming for a 1TB 3.5" drive and 20GB microdrives for handheld units.

What the Heck's Podcasting?

You've probably heard about this one: Yet another Pew Internet & American Life poll of internet users asking a clear question about each of eight terms: "Please tell me if you have a good idea what the term means, or if you aren't really sure what it means?"

Only 9% of respondents said they had a good idea what RSS feeds were, while 13% thought they knew the meaning of podcasting. *Truly* unfortunately, only 29% knew about phishing! Even worse, only 35% of those with broadband connections thought they knew what phishing was—and 12% of broadband users didn't know what a firewall is. Now there's a frightening number. (This all assumes that the sample is truly meaningful. Of numbers called for the survey, 35% resulted in completed surveys; make of that what you will.)

Art on the Cell Phone

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

museum, and includes some notes about Nokia initiatives, which include "limited editions" that you can pay to download to your phone.

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

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

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

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

Will the Internet Put Public Libraries Out of Business?

That's "just one question" from *Wired*, asked of Michael Gorman, Jessamyn West, and Sue Davidson of the Internet Public Library. I could grump about Davidson's answer, which seems to ignore the ongoing importance of *book* collections to say that "their role" will be "managed, vetted storehouses of organized data" with librarians "responsible for licensing, interpreting, and archiving information." (I'm reminded of why I've always disliked the name of IPL…and sometimes the assumptions of its people.)

Then there's the answer from Jessamyn West:

> I can see it happening. The people who can afford to buy computers and Internet service often stay home to read and do their research, which means that libraries are increasingly becoming places where poor people go to use public services. Meanwhile, communities all over the country have to make tough decisions about what should receive funding. When it comes down to a choice between putting money toward the police department and the library, there's no question which one has priority.

Arggh. The suggestion that all but the poor are abandoning public libraries is pretty horrendous. Jessamyn says she got a slightly different emailed question, "Do we still need libraries in a digital age?" to which she answered (in part):

> Yes.
>
> Is your question really "Do we still need books in a digital age?" in which case, the answer is more complicated, though ultimately yes.
>
> I guess my question for you is "Whose digital age?" because where I work, at public libraries in Central Vermont, the digital age is unfolding much more slowly and to much less fanfare than it is elsewhere… Libraries and librarians help people not get left behind by technology, by democracy, and by people who think that libraries and technology can't coexist and thrive symbiotically.
>
> We need libraries in any age, they're the human scale measurement for the information age.

She followed this up with a ten-minute phone call resulting in the first paragraph—and she doesn't claim that *Wired* actually misquoted her. While this doesn't say great things about *Wired*'s reporting practices, it's also unfortunate that Jessamyn was unwilling to offer a straightforward "Yes, books are still going to be around, as are libraries." And add that they will continue to be important public spaces used by all kinds of people, even as they're important safety nets. If only poor people use libraries in Central Vermont, something is very wrong…

Top 10 Web Fads

That's the title of a *CNET.com* item by Molly Wood, one of several "Power of 10" features to celebrate the site's 10th anniversary. It's an interesting list; you can

go there for details, or just consider how many of these you either visited, heard about, or cared about. (I'm a dud on fads; other than #9 as a reader, I score zero for "visited" and "cared about," although I certainly heard about all but Ellen Feiss.)

Hampsterdance (1998); Mahir (1999); All your base are belong to us (1998-2001); Dancing baby (1997); Hot or not (2000); Friendster (2003); Ellen Feiss (2002); Star Wars Kid (2002); Blogger (1999); JibJab (2004).

Well? How much of a web fashionista have you been over the past seven years?

Overlap among Search Engines

Here's another study where the results need to be interpreted fairly carefully—although it doesn't carry the scent of the rushed-out-the-door "Google's still bigger than Yahoo!" study. I saw it in the form of an August 2, 2005 *Internetnews.com* article by Susan Kuchinskas: "Dogpile: Search engines don't have much in common." This is PR country, given this lead paragraph's relevance to the headline:

> Dogpile.com, owned by Infospace.com, announced it added results from MSN Search to its meta-search service.

This is "backed up" with the results of a study done at Pittsburgh and Pennsylvania State "showing a surprising lack of duplication in the *top* results of the major search engines" [emphasis added] (where "major" means Yahoo!, Google, MSN and Ask Jeeves—in other words, the four used by Dogpile).

> When the researchers ran 12,570 different queries through search engines at Yahoo, Google, MSN and Ask Jeeves, they found that only 1.1 percent of the results appeared on all four engines, while 84.9 percent of the top results were unique to one engine. Only 2.6 percent of the results where shared by three search providers and 11.4 percent were delivered by two search engines.

> The researchers examined both paid and natural search results, but they tabulated only the results on the first page.

That first paragraph needs rewriting: "only 1.1 percent of the results appeared within the [*top 25? top 10? What constitutes the "first page"?*] on all four engines," and so on. It isn't true (or claimed) that 84.9% of results are unique to one engine; it *is* true that "relevance" algorithms vary widely. The spin from Infospace is based on this and the claim that "most

people never go beyond page one." Interestingly, Yahoo! had the highest percentage of unique first-page results (71.2%), Google the lowest (66.4%).

The claim is that Dogpile can somehow combine these four disparate first pages to produce the *best* first page—"based on consumer clickthroughs in the past" and at overlap in results. Dogpile *intermingles* paid and "organic" results, a true boon to merchants.

I'm surprised that a university research project studied "lack of overlap in sponsored links"—surely that's entirely a commercial issue? Maybe not. Oh, and there's the lovely final paragraph, in case you're in any doubt as to the scholarly nature of this piece:

> Dogpile.com also worked with Web traffic analytics firm comScore Media Metrix to determine searchers' success, based on the number of times users clicked on a link. ComScore found that only slightly more than half of all Web searches resulted in clicks on a link on the first page. Dogpile.com did better, garnering first-page clicks 63 percent of the time.

Are your clicks being monitored? Do you *expect* different search engines to order results in the same way? Perhaps more to the point, do you turn to Dogpile as a first choice in searching—and are you more likely to do so with the knowledge that it treats sponsored links as "first-page results"?

Quicker Takes

The July 2005 *PC Magazine* includes a surprising discussion of "undiscovered Office extras"—a *lot* of extra software tucked away in MS Office 2003, including a picture manager, built-in document scanning (OCR right in Office!), a clip organizer with a collection of clip art, and more.

➤ Not to pick on the June 2005 *EContent*, but Kinley Levack makes one humongous assumption in the lead to "HopStop.com: Mastering mass transit." Levack says, "New Yorkers always look like they know where they're going, barreling down avenues, cell phone in hand. *The secret is that they don't*, or rather they haven't—until now." [Emphasis added.] Right. New Yorkers are all lost but walking fast so we don't notice. *Sure* they are. I know *I'm* always lost in the city where I live and work unless I've got a cell phone or PDA feeding me directions!

➤ In February, I mentioned a skeptical commentary on multitasking (and left a word out in the process), suggesting that multitasking—or

"continuous partial attention," if you prefer—is a way to do several things badly instead of doing one thing well. A Johns Hopkins study appearing in *The Journal of Neuroscience* offers a specific reason this might be so: Evidence that "attention is strictly limited—a zero-sum game." The example used is talking on a cell phone while driving (and I see too many counterexamples to buy the idea that people talking on the phone *aren't* distracted drivers), but the test involved recording brain activity while students watched a computer display with "rapidly changing display of multiple letters and digits" while listening to three voices speaking letters and digits. "They found that when the subjects directed their attention to visual tasks, the auditory parts of their brain recorded decreased activity, and vice versa." They also found, *to their surprise*, that parts of the brain previously thought to be involved only in visual functions were affected when a participant was asked to pay more attention to the voices. The short finding: "When attention is focused on listening, vision is affected even at very early stages of visual perception."

➢ In the bad old days of "Cheap shots," I might have taken on David Bollier's "Why online commons are besting the mainstream media," posted July 19 at onthecommons.org. Bollier is excited about net media "empowering individuals" and "out-competing the market!"—and he just loves oversimplifications and universals. He seems to suggest that podcasting *only* works with Apple iPods—but that's OK, "as iPods become ubiquitous." Wow. Ten million in a country of 300 million is "ubiquitous." He says most TV is "banal, insipid and uninspired" *because* "our primary means of communication" have been turned over to a handful of people—where, presumably, we'll get *much* better-crafted TV and the like when everyone's doing it. (Surely you note that the average blog is better-written than a Stephen King or Gene Wolfe novel, don't you?) Finally, naturally, "we are forcing the mass media to re-tool its business models in order to compete with the strange new forms of non-market value-creation." (This post is full of exclamation points and some sense that

commercialization is still "free" and "the commons" as long as it's not Big Media. Or maybe I misunderstand.)

➢ Independent DVD replicators are upset with MPEG LA, the agency that handles royalties for DVDs, and are claiming that the companies in MPEG LA are violating European Union law. Why? Not because the royalty fees to produce DVDs have gone up. They haven't: They've gone *down* from $0.166 in the late 1990s to around $0.11 today. So what's the beef? DVD fabricators could charge studios about $1 per disc in the late 1990s—but the market's so competitive that they can now only charge about $0.30. Instead of paying 16.6% royalty, they're paying 37%, and they say that's not fair and reasonable. It is, of course, how patents work. The patent holder is not inherently required to account for the fact that licensors have entered an increasingly competitive market; indeed, that competitive market affirms that the patent holder is behaving in a fair, reasonable, non-discriminatory manner.

Followups & Feedback

Investigating the Biblioblogosphere

Surprise, surprise: People reacted to the NET MEDIA PERSPECTIVE that took up half of *Cites & Insights* 5:10!

Actually, I was surprised—by the breadth of the response and its *almost* entirely positive nature.

I won't repeat all the feedback: that would take more room than the original essay. If some excerpts seem to be non-sequiturs, that's because I've broken things down into categories.

Corrections and Clarifications

First, an important clarification that was in the original essay and disappeared during editing (my own editing—I have no one else to blame). I only considered *blogs written in English* for this survey, because I couldn't see how to do the metrics in blogs that I couldn't read.

I should also clarify "library people." They may not be librarians, and their weblogs might not be

about libraries, but they're associated with libraries and the blogs are in one of the directories I used.

Corrections and clarifications from bloggers mentioned in the survey:

➢ *Scholarly Electronic Publishing Weblog* didn't have an RSS feed until March 2005, although the blog began (as I noted) in June 2001. *The Laughing Librarian* began in March 2003. *Library Stuff* began on August 4, 2000, but the visible archives only go back to February 2003. *The Ten Thousand Year Blog* began in July 2002, again with some archives lost.

➢ Norma Bruce of *Collecting My Thoughts* was unhappy with the label "right-wing political." She prefers "political and personal with a strong voice." I urge you to read the whole discussion, including Bruce's concept of equivalent left-wing and right-wing personalities, at NBruce's journal on *LISNews*.

➢ *eclectic librarian* does have a good "About" page, which I missed.

➢ *Professional-Lurker* is written by *Lois* Ann Schmidt.

Walt at Random

First were comments attached to my pre-announcement of the essay and announcement of the issue, for example:

➢ Dorothea Salo (*Caveat Lector*): "I probably should link out more…"

➢ George Needham (part of *It's all good*) found it fascinating. "Greg" was impressed by the work involved.

➢ Laura Crossett reacted to my dislike for "biblioblogosphere": "Goodness, what have you got against neologisms?" [and noted that Shakespeare and Milton coined words] She's right, of course.

My apology for not clarifying "English-only" and restatement of selection criteria drew a number of comments including an assertion (on another blog) that I should rely on bloggers' own logs as a readership measure. Two commenters hoped that I would repeat the study. Some other comments, in part:

➢ Molly at *h20boro*, Waterboro library's blog: "I think we meet all the criteria, and the h20boro lib blog has been around since 2000" (…later, that *h20boro* may be a hybrid library/personal blog)

➢ tangognat at *TangognaT*: "Even though log files might be more accurate, in my case they are filled so full of spam, it would take way too much work to wring any meaningful stats from them..."

➢ Mark: "Not all of us have access to our log files either…"

Then came "the hornet's nest round" of comments on other blogs; in this case, I'll repeat part of the post itself as well as portions of the comments.

I was expecting this round of reactions to the "biblioblogosphere" piece to happen first (before the positive reactions) and with more force–that's why I came close to abandoning the essay. But it's really hard to throw away 50 hours of work and 7,000 words, particularly when you find the results fascinating.

Now it's happening, on two levels:

Critiques of methodology and limits, including my lack of non-English blogs (an editing error, explained in a previous post), claims that I should be requesting and analyzing server logs from every library weblog, and others.

Posts, two of them long, thoughtful, and even eloquent, that assert that the article is harmful because rankings are pernicious.

What's also happening, to my delight, is bloggers pointing out specific library weblogs worth looking at and providing their reasons for suggesting a look. Blogrolls don't do that; blogrolls are just sets of links. (There's an overlap between the hornet's-nest posts and those recommending lesser-known blogs.).

I'm printing and collecting all of this stuff (sorry, but that just works better for me than trying to put it all together looking at words on dozens of different web pages). I really do plan to blog about other topics here (one other one today, if time permits). I'll keep collecting feedback, direct and indirect, and almost certainly put some of it into a C&I essay.

Do I take two long essays that consider the profiles to be harmful more seriously than, say, 20 short reactions that want to see me continue? Is it really true that in every online "community" those who aren't included in a list will automatically feel bad about themselves and denigrate their own blog? (I find that hard to believe, particularly based on the reactions I've gotten from people not profiled…) Are library bloggers really that thin-skinned or…dependent on the roar of the crowd?

Damned if I know.

Responses in part:

- Fiona: "I don't think there's anything inherently wrong with rankings, after all the subscriber numbers in Bloglines shattered the myth that no one knew who was most read some time ago… It's great to see people suggesting who else to read…"
- Filipino Librarian: "I certainly hope you will continue. The next time someone refers to "blog people" as if we're all illiterate, we can show them some good examples backed up by numbers."
- Charles Bailey: "Hit a nerve with that article? Given that blogging is a very personal medium, that's what I would have suspected. But, it was a great article, and the frenzy of comment highlights its importance. Keep at it. I think you are on to something here."
- Angel: "I think the Filipino Librarian puts well in suggesting it will be work like yours that will validate what we do as well as show the rest of the world that bloggers are not an illiterate bunch…"
- Mark: "I think it was an interesting exercise in many ways, and while I clearly noticed and paid attention to your disclaimer about it being 'a' list, I am ambivalent about it—agreeing with statements made on both sides…"

Other Weblogs

I'm not sure just how often the study was mentioned in other weblogs. I have 26 printouts here, *not* from 26 different weblogs. Searching "Investigating the biblioblogosphere" on the major search engines yields 237 (really 70-odd) at Google, 145 (really 56) at Yahoo!, 80 at MSN, and—ta da!—16 at Ask Jeeves. "biblioblogosphere crawford" (without quotes) yields 322 (really 79) at Google, 172 (really 74) at Yahoo!, 664 (not tested for reality) at MSN, and a whopping 29 at Ask Jeeves. My guess is there have been 30 to 50 *distinct* discussions.

Favorable to Neutral Comments

LJ's blog entry (posted by Lori Bell) called it "fascinating and detailed." *diglet* called it a "very good environmental scan" and explicitly cited my note that it's "not a ranking." *scitech library question* called it a "lengthy and somewhat mathematical analysis." *Christina's LIS rant* said I was "using many of the proposed Paris metrics" and called it "worth a read even if you're tired of meta-blogging navel gazing." *tinfoil +*

raccoon honors me by calling the study "a terrific service." *librarian.net* called it "very fascinating" and called the numbers "somewhat subjectively chosen," and also noted that you can download my data and draw your own conclusions. *digitizationblog* called it "worth reading…if you want a broad survey of the variety of library blogs out there."

Caveat Lector said I was "far kinder to CavLec and me than either deserves" and hoped that I would "repeat the experiment eventually, because the library world has a lot of stellar up-and-comers…" *eclectic librarian* was "thrilled to pieces" to be included, and planned to look at a bunch of the other blogs she hasn't been reading. *Librarian in Black* called the piece "amazing" and suggested it's "the single most comprehensive look at the library blogosphere to date," suggesting it "could serve as a wonderful introduction for library staff who are eager to read library-themed blogs, but don't know where to start." *Lethal Librarian* was one of those who used the piece as a springboard to describe two less popular library blogs worth reading. *Professional-Lurker* found the study "interesting," noted the areas in which her blog stands out, and was a bit surprised to be listed as an essayist.

It's all good's Alane asked where I find the time (a dangerous question), had some fun with it, and ended with the kind of response I was hoping for:

> …I read many of the blogs with smaller readership because I like the topics, and/or the voice, and/or I know the author… My personal view is that blogging has brought to life a whole range of voices to our profession and that we are all richer for that. I've been working in libraries since 1975 and many of the names attached to articles or books were just that, names only. In the past few years, blogging and podcasting [have] given us all the platforms to broadcast widely, informally, and in the true voices of ourselves, the authors.

TangognaT is "not sure how I feel about lists of blogs. I think it is tough to rank blogs, just because there isn't a very reliable way of really calculating readership for a blog." She found the study interesting, especially because I detailed my methodology and provided raw data—and loved being ranked high for conversational intensity. "Yay for comments!" *Blisspix* discussed it at some length over three different posts. She (Fiona Bradley) calls for better tools to measure the blogosphere, not unreasonably calling my methods "convoluted." She notes that analysis of log files might be better—but "even they are wildly

inaccurate at times" and they're simply not available to those using hosted services.

Biblioblatherblog called it "fairly interesting," worried about adding too many new blogs to her "ever-burgeoning Bloglines subscriptions," and wished I could have done an age breakdown, "though I know it would be fairly impossible." She sometimes feels "like Methuselah in biblioblogland" and wonders whether she's "a voice of the over-forty blogger." (I believe *at least* six of the bloggers featured in the piece are over 50 and at least three or four are over 40.)

Wanderings of a student librarian offered a thoughtful perspective, quoting *Information Wants to be Free* as to whether library bloggers are just talking to each other and noting that the only way to find out is a survey. "Anyone looking for a research project?" She also wonders how many of the links to library blogs are from blogs outside the field. As one who wasn't in the group of 60 for which I did full metrics, she wasn't discouraged:

> I am gratified that my Bloglines subscriber number raises fairly consistently. I am thrilled whenever my Technorati or other ego feeds pull up a link to my blog… Email responses are great, too… Those things are motivating. A measure of broad reach, even if it were done once a year, really wouldn't inspire my writing from day to day.

The Information Literacy Land of Confusion notes that I didn't look at Google PageRank. He quotes Google's description of PageRank and offers my "short list" (the 60 fully analyzed blogs) in descending PageRank order (you'll find the full spot at lorenzen.blogspot.com, dated August 23, 2005). It's almost a pure bell curve, with one blog each at PR8 and PR4, eight at PR7, 16 at PR5—and 34 at PR6. *Library Stuff* quoted portions of this study, notes that he hasn't looked at his PR in two years or so, and admits (boasts?) that "My eyes glaze over when I read countless stats (probably one of the reasons why I have yet to read Walt's study in full)." He notes the PR list as "another indication of our obsession with statistics."

Negative Comments

Three of the longest and most thoughtful comments on INVESTIGATING THE BIBLIOBLOGOSPHERE were also the only negative comments (at least among those I saw). I referred to these as "the hornet's-nest round," and I'd expected heated criticism earlier on. I'll provide the base URL in each case because these are long

comments and I'm only offering brief (possibly biased) excerpts.

Information Wants to be Free (meredith.wolfwater.com/wordpress/index.php/) posted "Loving the long tail" on August 15. "I hate to play devil's advocate but I can't pretend that I think ranking blogs is a good thing." She found the study interesting "in an academic sense," then finds most of my metrics wanting: Bloglines is "notoriously inaccurate," Wordpress blogs "seem to be ranked more highly in web searches than other blogs," and rankings change daily, "so it's hard to depend on them as reliable measures of 'reach.'" About the only metric she finds worthwhile is "conversational intensity."

> Walt developed his own interesting system of ranking the biblioblogosphere. I don't know if I'd do it the same way. Actually, I wouldn't do it at all.

She goes on to explain why. It's an interesting discussion. "I don't like a popularity contest and I don't really see the need/point of ranking library blogs… Why perpetuate the insular/clique-ish stereotype of the biblioblogosphere by actually ranking them and leaving certain people out in the cold?" She imagines that some bloggers not in the fully analyzed set "found this to be discouraging—though I hope they didn't care… A lot of my favorite library blogs were not in Walt's Top 50…. How many people reading *Cites & Insights* weren't already aware of most of those Top 50 blogs?" She goes on to describe eight worthwhile blogs that weren't in the fully described set.

I could take issue with elements of Meredith's essay. I *do* take issue with the constant repetition of "ranking" and "Walt's Top 50." The article does not provide a numbered ranking. It explicitly says "*a* top 50," one of many possibilities (there are 60 blogs in the list, an odd way to prepare "Walt's Top 50"). I did not claim these were the most popular blogs, the most important, or anything of the sort. I claimed *and do claim* that they are 60 non-group, non-official-library blogs that have demonstrated reach. Nor, frankly, am I terribly worried that some poor soul will be so disturbed by being "left out in the cold," and will take this study so seriously, that they will be "discouraged." *If they are, then they're writing a blog for the wrong reasons.* After another response (noted above) wondered about age breakdown, this blogger chose to do her own survey of library blogs. I participated, and look forward to the results.

Random Access Mazar (www.mazar.ca) posted "The A-List and the Z-List," also on August 15, finding my essay "depressing" and asserting that *any* sort of ranking "invariably causes hurt feelings, conflict, and disappointment, and even results in some undue criticism being levelled at the chosen ones… Ranking people cheapens the whole process." She says "without reservation that someone somewhere felt hurt by being left off his ranking."

> For me, the point of blogging, and the joy of blogging, is in having a place to write things down. For me writing is thinking, and I love to be able to share my thoughts with anyone who's interested. Rankings therefore don't bother me much, because my goal has never been to please other people.

So Mazar isn't at all unhappy that *she* wasn't on the list of blogs with full metrics, but she's 100% certain that other people *are* unhappy that they weren't included. This discussion continued in email, in which she said explicitly "some of us will always feel uncomfortable with anything that smacks of ranking." And, elsewhere, "As soon as there's some sort of prize to be won, the race looks a little different, you know?"

explodedlibrary.info posted an August 17 comment (www.explodedlibrary.info) that isn't an attack, but she does have "very mixed feelings." She blogs for herself, "and will continue to do so irrespective of rankings or recognition." But she knows there are different reasons for blogging, and "who are we to say that there are right and wrong reasons for blogging?" She does think "it's possible to say that blogging for fame/ego is not advisable because it's such a fickle game." She would have been more interested in "Walt's purely subjective 60 favourite blogs" and asserts, "every objective appearing calculation also contains subjective judgements/assumptions, sometimes they're just more hidden."

I don't have a list of 60 favorite blogs, and I can only say that I offer the actual numbers used and the reasons for using them—and that I would have seen to it that at least one blog did not appear on that list if I was doing it subjectively. (Two, actually: *Walt at random* wouldn't have been there.)

She adds a quick response to Rochelle Mazar's dislike of any ranking: "otoh why are some people so uncomfortable with rankings, because ranking is ubiquitous on the web, as in Google pagerank and Technorati etc, etc."

That's a version of my response to claims that no rankings should ever be done. Life doesn't work that way. If you don't want people to feel "lessened" because they come out shorter on some measurement, then just within the blogosphere you'd have to:

➢ Eliminate Bloglines subscriber counts, since as soon as you sign up for your own Bloglines feed, you realize that you will probably never have as many subscribers as Jenny Levine.

➢ Eliminate blogrolls. When you list a bunch of other blogs without comment, you are implicitly saying that blogs not on the list are not as worthy as those on the list. (*Random Access Mazar* has a blogroll with 20-odd names.)

➢ Eliminate Technorati, Blogpulse, Feedster…

➢ Eliminate search engines. Who hasn't done an ego search? And if you ever search another blog name, well…

➢ Eliminate articles that discuss blogs and name any of them, since any such naming points out certain blogs as specifically worth noting.

In the rest of the world, you'd have to eliminate grades, degrees, *Who's Who*, competitive sports (competitive *anything*, actually), awards, honors…the list goes on and on.

If you regard blogging as a race and being on a list as "the prize," then you've already lost. If failing to be included in one study of one aspect of blogs by one commentator who's not on anyone's A-list causes hurt and disappointment, that's a shame.

Ten bloggers who weren't in the set of blogs receiving full metrics chose to comment on the study. Eight of the ten *liked* the study; several of them want to see it done again. Even calling *explodedlibrary.info* negative, which may overstate her criticisms, two of the ten were unhappy. I can live with that.

Email Feedback

Most of this has already been discussed.

Marlène Delhaye of *Biblioacid*, a high-profile French library blog, had raised the question of my study's lack of non-English blogs; after I explained it, she sent a kind note—and noted that most of the metrics (except the wholly subjective "voice") can be applied to any blog, regardless of language. She's right.

Von Totanes, *Filipino Librarian*, had some kind words and asked whether anyone had suggested using Google PageRank. Since the posts using PageRank on the fully-analyzed list hadn't appeared yet, my re-

sponse was "well, now you have…," and I also responded that a followup probably wouldn't use Google link: A further interchange clarified *why* I'm likely to drop that measure: Google *admits* that link: returns only a sample of the actual results, making it useless as a metric.

"If There is a Next Time"

In several of my responses to email and comments-on-comments, I prefaced possible changes in methodology with "ITIANT"—If There Is A Next Time. I gathered as many of the responses as possible.

I don't need to make a decision until next June or July, since, ITIANT, I'd use either March-May or April-June for analysis of posts. It's possible that more feedback will change my mind, or that other circumstances ("life happens") will make the decision for me.

At this writing, I believe a similar study *is* likely to be repeated with changes. Likely changes so far, with more possible after some analysis in mid-2006:

➢ Non-English blogs listed in one of the key directories *will* be included in the broad analysis and, as feasible, in full metrics. I can't say much about "voice" or personal/professional/political balance when I can't read the posts, but that description is a tiny part of the study.

➢ Google "link:" probably *won't* be used as one of the raw measures unless Google changes its practice. Some other raw measures may be added. Google PageRank *might* be part of the raw measure (or of extended measures), but only part. No single measure will play a dominant role in the analysis.

➢ Depending on the energy and time I have available for a repeat study, I might issue an open invitation for "blogs not in one of the directories" to send me names and URLs, capping the total at some level that's plausible (250? 300? 400? Nah, probably not 400…).

➢ "Not official library/organizational blogs" will almost certainly remain as a filter. "Not group blogs" will probably be modified to allow blogs with a large handful of contributors that retain some overall coherence—if I can find an objective line that excludes (for example) the LITA and PLA blogs but includes (for example) *Resource Shelf* and *Open Access News*. I think that's possible.

➢ I will not make any effort to note demographics of bloggers other than their names.

➢ The spreadsheets will be made available, which means there's *no* way to prevent someone who wants to come up with a "ranking number" from doing so. I can't be transparent about methodology and obscure about results: The world doesn't work that way.

➢ There won't be a published numbered ranking (just as there wasn't this time), but there will be ordered lists of blogs standing out on individual metrics.

➢ I might try doing the metrics on a bigger chunk of the overall list, if time allows, possibly by simplifying one or two metrics.

➢ I'll continue to pay attention to suggestions but with no promise they'll be followed.

➢ Those who oppose comparisons of any sort will be unhappy. I can live with that.

Studying the Blogosphere: Other Voices

I don't believe there is a true "A list" in the biblioblogosphere. Many people sure do believe there's an "A list" for the blogosphere in general and that being on the A list is meaningful. So, for example:

➢ *j's scratchpad* notes that Feedster now has a "top 500" list—and provides a graphic you can put on your blog to brag about being a Feedster Top 500 member.

➢ *Library clips* has an August 12, 2005 essay (5 print pages), "Blog ranking: incoming links??" The blogger notes some "hot lists" and their methodologies and discusses problems. It's all about finding lots of metrics so you can find the "hottest blogs" for your own definition of "hot." I'm more interested in finding the bloggers whose posts inform me and make me think: That's not a quantifiable metric.

➢ *TechBlog* also has an August 12, 2005 essay (3 print pages), "Blogging for its own sake," which raises the question of why this person blogs and why other people blog. *TechBlog's* author doesn't feel they get "enough readership" and offers strategies for increasing readership. It's an odd post with fancy numbers and includes questions I can't imagine asking of my own blog—but then, I'm *not* out to score a huge readership, just an interested one. ("Fancy numbers"—one section includes

four equations and requires quantifying nine separate factors.)

➢ *Science Library Pad* (August 21, 2005) discuses blog ranking and metrics, including projects from IceRocket and Feedster.

➢ *Napsterization* has a six-page August 6, 2005 post, "Link love lost…" (it's a *long* post title) that includes a table with *nineteen different metrics* for evaluating a blog, looking toward a "community-based algorithm." Here, as with so many other discussions, "authority" seems to be the target. I'd argue that no set of metrics will measure authority in blogs any more than it would in, say, magazines or books. I find the essay bewildering. "Part of what we want is a rich user generated ontology resulting in topic groups that is constantly adjusting to find what's delightful, useful, interesting, across blogs."

➢ *thejasoncalacanisweblog* (part of Weblogs, Inc.) has an August 2, 2005 "Blog 500 Challenge" in which Calacanis offers $10,000 in cash if a programmer builds a "500 list that kicks butt" for Weblogs, Inc., or $50,000 in advertising to the first person to come up with such a list. Such a list would use links rather than number of inlinking sites, look at the most recent 12 months rather than the full history of links, have a separate sublist for "up and comers," and be updated constantly. Lots of comments…relatively few of them along the lines of "who needs these lists?" Weblogs, Inc. is in the advertising-on-blogs business, as the blog's page clearly shows, and if that's your game, "hot lists" are probably part of it.

If there is a next time, it will once again be a snapshot looking for *specific* areas of interest: Which reasonably widely read library-related blogs are composed of lengthy essays? Which post frequently? Which have loads of comments relative to the number of posts? The thought that a metric, no matter how complex, could predict "what's delightful, useful, interesting across blogs" doesn't work for either my "numerate" side or my humanities core. I don't believe you can quantify meaning or delight or interest any more than you can love or friendship.

Non-Biblioblogosphere Feedback

Finally, a couple of items that have no relationship to INVESTIGATING THE BIBLIOBLOGOSPHERE.

Seth Finkelstein noted that my comment about liability for a handgun company that advertises its products as "Perfect for taking out your old lady" (July/August) wasn't all that hypothetical. One gunmaker apparently did design its products and marketing strategy in ways that may seem questionable (advertising an assault weapon as having excellent resistance to fingerprints, for example, or specifically designing a gun to accommodate silencers or to be shot from the hip). See *Infothought* for August 9 (sethf.com/infothought/blog/).

Scott Pope had a detailed reaction to my "current credo" (in the July/August issue). Here it is, in full:

> When reading Walt's latest *Cites & Insights*, I found myself thinking too many of his points sound good but don't fit the reality of librarianship, as I normally feel when I read his newsletter. Don't get me wrong—I love reading his newsletters because he writes in a serious way about serious things, but his conclusions usually sound too academic by far for my taste.

> What follows is my response to each point of what he calls "my current credo" in the latest newsletter:

> 1. "Good public and academic libraries are both physical institutions and sets of services. They serve a variety of purposes within real communities and colleges, and some of those purposes can only be served effectively through physical libraries."

> My response: Bad libraries are these things too. What sets good libraries apart is that the "physical" buildings are appealing and the "services" become desired.

> 2. "We will continue to see revolutionary predictions based on oversimplification, bad economics, infatuation with technology, and failure to appreciate people. Librarians who fall prey to such predictions will suffer, as will their users. Librarians and library supporters must be ready to challenge unlikely projections, analyze faulty economics, and assert the need for choice and the importance of both history and the present."

> My response: "Revolutionary predictions" may not be a bad thing. They may cause an individual library to publicize a service in a new way or adjust services so they match with already publicized projections. In other words, even if the projections are "unlikely," they might be a selling point. For example, you can use Google (like the University of Texas uses Google

Scholar) and benefit from all the buzz even if it is overblown by far.

3. "Technology and media will continue to interact in unexpected ways, but ways that will lead to more rather than fewer media. Different media serve different kinds of stories well, and new media should enable new kinds of stories—but the kinds of stories that books serve continue to be critically important for libraries."

My response: "More" media isn't always better. For example, I'd love to get rid of a lot of our microfilm. Some media will waste the time of the reader/learner/experiencer.

4. "Print books will survive, and will continue to be at the core of all good public libraries and the humanities and social science portions of good academic libraries."

My response: Print books will also continue to be at the "core" of hard science libraries.

5. "All libraries and librarians need to deal with increasing complexity, not as `transitional' issues but as the reality of today and tomorrow."

My response: Some great librarians will let others handle the "complexity," and concentrate on service, advocacy, salesmanship, and education. In other words, professionals concentrate on the abstract concepts such as helping students and fan off the web programming to a systems librarian.

6. "Libraries must serve users—but all users, not just today's primary users. There's a difference between being user-oriented and pandering, and it's a difference librarians should understand."

My response: Library school jargon. True, good libraries attempt to profile and reach out to non-users, but, in reality, libraries that truly serve "all users" have an extremely small user pool.

7. "Libraries matter, and librarians should build from strength. There are many fine public and academic libraries and many more that do remarkable work with inadequate resources. The goal should be to improve and diversify from what libraries do well, not to abandon existing services and collections in search of some monolithic futures, whether all-digital or otherwise."

My response: Some libraries, such as bad libraries, might be better off if they "abandon existing services and collections." While building on your strengths may sound good, it may prove fatal. Let's face it, weeding is not done enough in a lot of libraries that are seriously decaying.

8. "Libraries will change, just as they have been changing for decades. Good libraries will maintain live mission statements—and the missions won't change rapidly."

My response: A lot of the changes have been too minor. A lot of libraries look like the only changes they have done in decades is adding a computer area. Good libraries will change more than technology. Few mission statements reflect the reasons the customers actually go to the library, so maybe it would be good to change those mission statements.

9. "Effective libraries build communities, and the need and desire for real communities will continue to grow. Libraries that work with their communities should prosper; those that ignore their communities will shrivel."

My response: Hopefully libraries will do more than "work with their communities," like stun the heck out of them, or inspire them, or even make the community grow because of their existence.

I'm inclined not to respond, at least not here, not yet—not because I think my "credo" is wrong or weakened by his responses but because I think his responses are useful additions. I will make one point about #6 (where on earth would I pick up "library school jargon"?): I thought that *within your potential user base* was implicit in "all users." A good academic library should be designed to serve the next generation of students and faculty, not just those currently in residence. I think there are definitional issues elsewhere as well, but will leave those to the reader.

URL: cites.boisestate.edu/civ5i11.pdf