Cites & Insights

Crawford at Large

Volume 3, Number 14: December 2003 ISSN 1534-0937 Walt Crawford

Bibs & Blather

The Stuff and Nonsense Issue

As promised, the Stuff is back—*lots* of it (around 8,000 words) as I start cutting 24,000 words down to size. I want to start out fresh for the new year—and this issue almost came out at 26 pages. With considerable reluctance, I cut a 4,000-word piece that's a bit less relevant to library issues, and trimmed another 4,000 words here and there. I suppose eliminating one-third of the work you've done is good for the soul…

Nonsense? I'm inclined to say that the copyright hearing notes and the Ashcroft brouhaha are both heavy on nonsense—but not innocent nonsense, sad to say. I count feedback and the "Carolinas" commentary as Stuff, just as I do the trends section. As for Bibs & Blather: Maybe both.

Volume 3: Not Done Quite Yet

If you bind *Cites & Insights*—and in my egomaniacal moments, I believe that at least 49 libraries in North America *should* have bound sets—don't send off the issues just yet. Once again, I'll be issuing a volume index (and volume title sheet), some time between now and Christmas. Best guess: Very early December. Maybe sooner.

Where's the Reader Survey?

I'm not doing a formal survey this year—not because I don't value your opinions, but mostly because I didn't get around to designing the kind of survey I'd want to do. I'm also less convinced that major planned changes in *C&I* would be needed or likely for 2004.

But since this is the last regular issue for the volume, I would like to issue a *special* invitation for your feedback, in addition to the standing invitation. If you're inclined, send me email (user wcc, domain notes.rlg.org) answering any or all of the following questions in whatever manner you choose:

- What continuing aspect of *Cites & Insights* do you find most interesting or most valuable?
- What continuing aspect of *Cites & Insights* do you find least interesting or least valuable?
- What would you most like to see me do here that you believe I could be good at?
- If there's something that bugs you, let me know—*with the exception of PDF*.

"Continuing aspect" can mean ongoing topical coverage, the typical length of items, writing style, or whatever. If I'm injecting too much personality for your taste, you can grump about that: I promise not to be offended, although I also promise to keep this a zine, not an objective journal. If you'd like to see *more* "off-topic" stuff, let me know that.

Inside This Issue

And Still More Blather

Alert readers of *American Libraries*, and particularly those who remember a "sharp things protective act" (*C&I* 1:7) might be waiting for me to rev up Cheap Shots & Commentary so I can bash Will Manley's September 2003 "Will's World," wherein he opines that librarians shouldn't make the crucial CIPA-related decisions in public libraries. He says those decisions should be addressed by library boards of trustees. "In a nutshell, trustees make library policy and librarians put that policy into practice."

If a public library's charter is such that the trustees set library policy (as I assume it is in most cases), then I have no argument with Manley (on this point!). It's imperative in such cases that librarians be effective communicators, so that they can help the trustees see the full set of issues involved. When Manley goes on to say, "Each community is different and each public library should reflect those differences"—well, since I've been saying much the same thing in my keynotes for years, it would be a bit churlish to disagree. So I won't.

Perspective

In Your Jets I'm going to Carolina

Sorry about that, and apologies to James Taylor. It was an odd coincidence that my only travel this fall was to Winston-Salem, NC in September and Charleston, SC in November: the North Carolina Library Association and the Charleston Conference. These two Carolinian conferences have two things in common besides geography: They involve lots of librarians and they're both worthwhile. I did a plenary session (the Ogilvie Lecture) and a "table talk" at NCLA; I was at the Charleston Conference to listen and learn, but was also on one panel.

Impressions of NCLA

"Libraries—A North Carolina Value: Enriching – Inclusive – Essential." That was the theme for NCLA's 55th *biennial* conference, September 23-26 in Winston-Salem. Biennial? Yep: This is that rare state conference held every other year.

I was the only plenary speaker from out of state—and, for that matter, the only one who wasn't a North Carolina author. NCLA doesn't have a big awards banquet, or for that matter *any* dinner events. There were two group luncheon events on Wednesday, three breakfasts and *five* topical lunches on Thursday, and three more lunches on Friday. All-conference events included an opening general session Wednesday morning (featuring Allan Gurganus, author of *Oldest Living Confederate Widow Tells All);* a vendor reception that day; my speech Thursday afternoon; a Thursday evening all-conference reception at the Forsyth County Public Library; and two unusual Friday events: an all-conference *breakfast* (8 to 10 a.m.) featuring North Carolina poets and a dessert reception featuring Doug Marlette, creator of the comic strip "Kudzu" and a long-time editorial cartoonist. Every state conference (that I've attended!) leaves lots of open slots for exhibits; this was no exception, with a total of eight dedicated time slots for exhibits.

I wasn't able to attend many concurrent sessions. One on enhancing library catalogs offered interesting descriptions of various enhancement systems (adding tables of contents, etc.) but also some talks that came off as sales pitches. A session on the future of the book included a substantial overestimate of the market share of ebooks but also offered a good range of thoughtful perspectives, including publishers and booksellers as well as librarians. A bookseller noted that "kids these days" are buying *plenty* of books and that there's no letup in demand for print. A publisher discussed a specialized ebook created three years ago—that has sold all of *three copies* since then. The panel and active audience questions raised a number of issues, but (as usual) nothing was resolved. John Budd did a formal lecture on "how we know what we know," an interesting presentation that felt a little out of place given the 25 to 30 people that attended. Knowledge is a tricky thing—can you actually *know* anything you didn't personally test?

I was fascinated by Timothy Gotti's presentation, "Rise of the young turks: Generation X as managers in libraries," where this 31-year-old head of cataloging discussed his experience with 18 employees, most of them Baby Boomers. Bright, interesting, and informal, he was also thought provoking, if a little enthusiastic about technology for my taste.

Tim Bucknall of the University of North Carolina, Greensboro led a fine session on OpenURL—one that opened my eyes to the problems of OpenURL sources. I won't comment on specifics, but it's fair to say that I learned more about OpenURL issues from the resolver-builder's side than at any other session I've attended. (I've *used* dozens of resolvers, but always with Eureka OpenURLs; Bucknall gave startling examples of how OpenURLs arrive from some other sources.)

Interesting programs, but so many conflicts that some programs probably didn't get the audience they deserved. Great people and good events, a good conference facility attached to a pretty good hotel. They brought in several out-of-state speakers. North Carolina produces some excellent wines and many thoughtful librarians. I enjoyed the experience thoroughly.

Incidentally, if you're in Winston-Salem, Old Salem is worth half a day. It's where the Salem half of Winston-Salem began, as a Moravian settlement, and it's still mostly residences, but with a good sprinkling of authentic crafts and displays for a reasonable fee. That's also the site of the Museum of Early Southern Decorative Arts, definitely worth a visit, and the first museum I've been to that you can only see as part of a guided tour. My only negative comment: one-person "groups" should expect to be treated badly by the Old Salem Tavern restaurant.

Sketches of Charleston

Bring together 700 librarians, publishers, database providers, intermediaries, and consultants, all on the general theme of book and serial acquisitions and use. That's the Charleston Conference, a topical

conference with scores of speakers and a crowded schedule. It's a truly unusual conference: Always in the same place, always more-or-less at the same time (the first week of November, give or take), now in its 23rd year, and bringing together sometimes-"enemies" for frank, open discussion.

This conference has two advantages in trying to present an enormous range of content. First, the "exhibits" are confined to a six-hour session with one table for each exhibitor on Wednesday, ending in a wine-and-cheese reception. Second, Charleston starts early each day, ends late, and works on the basis that you're expected to get up and walk around (or leave) whenever you need to.

Thursday's first set of plenaries began at 7:55 a.m. and ran to 11 a.m. (with a 20-minute break at 10 a.m.); the final set of concurrent sessions ended at 5:15 but was followed by publisher/vendor forums until 6:30, before the 7 p.m. conference reception. Friday: 8 a.m. start, 6 p.m. finish, followed by an evening program. Saturday began with concurrent breakfast sessions from 8 to 8:55, and although the conference finale ended at 1:30 p.m., there was a rump session Saturday from 4 to 5:30 p.m.

"Oh, but there are lots of long breaks in those days, right?" Sure: One twenty-minute break each morning and afternoon (two Thursday morning). Otherwise, you had to skip sessions to have free time. I skipped one plenary session and the "lively lunches" Thursday and Friday, but that meant missing one of ten concurrent Thursday sessions and nine Friday sessions. But I still went to five plenaries (of six with a total of 35 speakers) and, including the Beastly Breakfasts (included in the registration price), six of seven concurrent slots with a total of 42 sessions and—well, I'm not willing to count up all the speakers.

David Levy opened with a keynote on "holding on to reality," talking about the need to be more real and take time to contemplate (among other things). Yet another "future of the book" panel followed, including yours truly. There were sessions on open access, ebooks, deep log analysis, user statistics and usage levels, collection development, license negotiation, government publishing, resource integration and OpenURL, the next generation of learning (a classic KTD speech, but others loved it), digital archiving, the integrity of full-text databases, the "Faxon fiasco," pricing models, censorship, copyright, database evaluation…

Tired yet? I'll stop. I tend to take very few notes at most conferences. I have 28 pages of hand-written notes for this one. If I'm offering less detail here than I did for NCLA, that's because there's so much it's a little overwhelming. The breaks are deep in discussion; I wound up having two business lunches by accident; the all-conference reception (at Charleston's acquarium) was great; and Heidi Hoerman from the University of South Carolina's School of Library and Information Science put out a fine single-page "News from Yesterday" that offered fascinating tidbits from the sessions. There will be a proceedings volume, probably right around the time of the 2004 conference. I won't say all the speakers were great (two plenary speeches left me completely cold, and others ranged from brilliant to boring), but I will say that—as with last year—I found this a lot more thought-provoking than most conferences.

The Good Stuff

Adams, Irene, "Overcoming web page printing problems," *Online* 27:5 (September/October 2003): 36-41.

If you've never had trouble printing out a web page, you're a lot luckier than I am—or you don't print web resources much. There's nothing like printing out a 20-page article, noting that you're missing about two words on each line, and finding there's no good way to solve the problem. There have been articles that I planned to use in *Cites & Insight* that never made it because the hassle of getting a readable print copy outweighed the value of the piece. Sometimes, I believe, web designers are deliberately preventing printing. I wonder whether the creators of MovableType just hate paper?

This **highly recommended** article won't solve all your printing problems—but it will surely help with some of them. "How long is a web page? As long as it wants to be." How *wide* is a web page? That depends—and some width problems are nearly insuperable. Printing framed pages is typically horrendous, but it's only one of the horror stories Adams tries to cope with.

"Many Web pages use white text on a colored background. This looks very nice, but presents a printing problem." I was going to take issue with "looks very nice," but if you don't care about readability (or you're only using a headline or a few key words), she may be right.

Adams has applied some ingenuity to solving various problems. Her tips won't help in all circumstances, but they can get you going.

Adams, Stephen, "Information quality, liability, and corrections," *Online* 27:5 (September/October 2003): 16-22.

An interesting and informed discussion of how (mis)information survives and what it can take to fix it. Adams discusses categories of errors and "inappropriate quality," with appropriate examples. Some examples are slightly astonishing. For example, one particular datum crucial to environmental studies, the octanol-water partition coefficient for certain pesticides, has been reported over 55 years with "up to 4 orders of magnitude variation with no convergence with time" for what should be identical data. (Four orders of magnitude: that's a difference on the order of 10,000 to 1.)

I could have done without a touch of casual and pointless racism that the British author is probably entirely unaware of—but the article as a whole is a valuable and generally well-written discussion of a topic we tend to ignore some times. It's easy to assert that Serious Published Literature represents higher quality than the web—but that's only true if the refereed stuff is legitimate and hasn't been undermined by later events.

Banks, Michael A., "Mining web data on a budget," *Online* 27:5 (September/October 2003): 32-5.

Here's a fascinating look at "out of the box" ways to do casual research on a budget—ways to extract free information from primarily-commercial databases legally, approaches to data resources you might not consider for a given topic. Worth a look.

Byers, Fred R., *Care and handling of CDs and DVDs: A guide for librarians and archivists*, Council on Library and Information Resources and National Institute of Standards and Technology, October 2003. ISBN 1-932326-04-9; $15 (or downloadable for free from www.clir.org or www.itl.nist.gov). 42 pages.

I've been waiting for something like this for quite a while; now that it's here, I **highly recommend** it. Unfortunately, it does *not* deal with the kind of damage likely in library circulation or the problem of hublock-related damage and how to avoid it. It does, however, offer lots of information about writable and rewritable discs, the physical aspects of optical discs, and what that means in terms of lifespan and casual damage. It begins with a one-page "quick reference guide for care and handling" that includes lots of useful recommendations, although one or two need additional context.

I was *not* aware that you shouldn't store discs flat for long periods of time (as you will in many CD holders), and it seems a bit counterintuitive. As for the advice not to use adhesive labels, it's complex (as the guide states): the labels could delaminate over time and interfere with disc drive operation (and attempting to remove a CD label is likely to damage the most vulnerable part of the disc), but such labels may *add* protection for discs not intended for long-term storage. I apply labels to the CD-Rs I burn, but I don't necessarily expect any of them to last more than five years.

Some items surprised me. For example, polycarbonate—which makes up most of a CD or DVD (it's the plastic layer or layers)—will absorb moisture to some extent, which can cause problems and may help explain why some users in tropical and subtropical areas report that DVDs and CDs don't last very long. The metal layer in most CD-Rs and DVD-Rs is silver or a silver alloy (some use gold)—but aluminum for rewritable discs and most prerecorded discs. While burned recordable discs should have a life expectancy of 100 to 200 years (according to the report), and even rewritables should last 25 years or more, *blank* recordables should be used within five to ten years of purchase.

Write-once discs can be damaged by prolonged exposure to sunlight—and *all* writable discs can be damaged by exposure to heat. That may explain why some of the CD-Rs I play in the car suffer from playing problems after a while, while that doesn't happen with pressed CDs. (The problems don't seem to affect anything except the car player, but now I know why! Incidentally, I find that CD-Rs using azo dye, the dark green or blue ones, work better than the more common silver phthalocyanine CD-Rs.)

There's a *lot* more in this well-written, clear report. You already know (don't you?) that you shouldn't write on the label side of a CD with anything other than a marker designed for CDs, and particularly not with solvent-based markers or fine-point ballpoints. The latter advice may be irrelevant for single-sided DVDs, since those have six-tenths of a millimeter of polycarbonate between the label and the information surface (unlike the very thin lacquer coating on CDs). There's a clear and correct glossary and an extensive bibliography.

Fallows, Deborah, *Spam: How it is hurting email and degrading life on the internet*, Pew Internet & American Life Project, October 22, 2003. (www.pewinternet.org). 43 pages.

"Spam is beginning to undermine the integrity of email and degrade life online." That's the first text in this report. "More people are reporting they trust the online environment less." I'd call that a silver lining in the spam cloud: People *should* be more wary and skeptical of online stuff, including email. "70% of email users say spam has made being online unpleasant or annoying." That figure seems a little

low, but it's possible that 30% of email users get all their email through ISPs such as AT&T, who seem to have very effective spam blockers—undoubtedly at the expense of losing legitimate email.

The report is chock-full of fascinating figures, and Fallows gets full credit for using broad percentages in line with actual sample sizes. 1,272 email users responded—and Fallows *never* cites a decimal point for a percentage. Since the margin of error is roughly 3%, that's sensible—but rare in my recent experience with survey-based reports.

If you're interested in how others perceive spam, I **recommend** the report. Just a few figures: A third of personal email users say at least 80% of their mail is spam—but only one in ten office email users say at least 60% of email is spam. And spam continues to be a problem partly because 7% of respondents actually ordered a product or service!

People who get lots of email have proportionally more trouble with spam than others. (Most workers don't really get that much email—44% receive 10 or fewer on a typical day.) Surprisingly, 32% of respondents considered commercial email *from senders that you've already done business with* to be spam, which I believe exaggerates the problem. (I have yet to see a supplier I've done business with that won't turn off unsolicited email as soon as you ask; of course, those are the only ones where I'll ask.) Most respondents are sensible: 86% delete spam immediately without opening it. Or maybe not: 67% clicked on "remove me," which tells the spammer that they have a real email address. And, sigh, 33% clicked to get more information—at which point I don't believe they're entitled to call it spam.

Good, Robin, "The future of RSS—Is e-mail publishing dead?" *LLRX.com*, September 29, 2003.

Robin Good, an alternative journalist/editor, is an RSS enthusiast. "RSS is good because it gives back to individual users the power to choose and select content." (I could argue that, at least to remain part of society as a whole, excessive use of that power is a bad thing: The broadening effects of scanning a daily newspaper are important. But I've written about that before and will again.)

The first subhead is "Is RSS going to replace e-mail?" But Good immediately responds, "The question has been wrongly posed." As a reader I find it awfully difficult not to stop there and say, "Wait a minute: *You* posed the question that way, and gave it as a heading." And the answer really is that "RSS may [be] well poised to substantially challenge e-mail in its ability to be the best and preferred distribution/subscription mechanism for newsletter publishers on the Internet." Whew. One sentence later, and I find an *assumption* that email has been the "best and preferred" mechanism, an assumption I'm not ready to accept on faith.

It gets better. "RSS is also something that once you have read its description, you know less about it than you did before." I'm not sure what an "eco-confused information space" is, but Good's explications of positive aspects of RSS-based news feeds are pretty good. Good offers 17 advantages in all, some of which are essentially redundant and a couple of which hit me as just plain odd. (Particularly the final one, "RSS paves the road for true ethical marketing"—well, you'll just have to read the paragraph. Maybe it makes more sense to you.)

The list of problems is shorter, 11 in all, some of which are killers for some newsletter/zine editors—e.g., limited formatting options, the loss of context (that isn't the way Good puts it), and the fact that RSS is more "one-way" than email.

Things get rough again a little later, where Good offers a "vision" in which authentication is provided for email—and you "block everything [in email] not coming from trusted or 'known' sources." Yes, I'm sick of spam—but not so sick that I'm willing to say that, if I don't already know you, there's no possible way you can send me email. I hope it doesn't come to that; if it does, email is essentially dead.

I'm not sure I can heartily recommend this, but it does offer concise thoughts on RSS for newsletter writers. **Worth reading** and thinking about.

Holt, Patricia, "10 mistakes writers don't see (but can easily fix when they do)," *Holt Uncensored* 376 (October 9, 2003). www.holtuncensored.com/members/

I still remember Patricia Holt as book editor for the San Francisco *Chronicle*, but that was long ago—before *Holt Uncensored* and her editorial consultancy. She's "concerned about the amount of time I've been spending on easy fixes that the author shouldn't have to pay for." She says this list—which, with comments, runs to seven print pages—could also be called "10 common problems that dismiss you as an amateur." (She uses ALL CAPS FOR THE TITLE, which is one of those online problems that dismiss you as an amateur. But never mind.)

The list? Repetition ("crutch words"); flat writing; empty adverbs; phony dialogue; no-good suffixes; "to be" words; lists; "show, don't tell"; awkward phrasing; and commas. She quotes published writers—sometimes important ones—as examples of each problem. "The point to the List above is that even the best writers make these mistakes, but you can't afford to. The way manuscripts

are thrown into the Rejection pile on the basis of early mistakes is a crime. Don't be a victim." (Hmm. Pointless Capitalization within a Sentence. But then, she repeats a three-word mantra, three times, in "empty adverbs," just two down from "repetition." In (rightly) objecting to overuse of the "be" words, she says, "If used *to* [sic] often, this crutch…"—but "failing to review your copy carefully" isn't listed as one of her mistakes. (Maybe there's a verb "often" that I'm unaware of.)

Okay, I'm poking fun at this **recommended** piece. When you set out a list of "10 ouchies" that are "so obvious to literary agents and editors" and quote examples of the mistakes from published works by Hilary Clinton, Sheldon Siegel, Julia Glass, Gail Sheehy, and Graham Greene, you're asking to have your own style picked apart. I'm going to stick this piece in my rarely-consulted "look at it again sometime" pile. Maybe some day I'll check my own writing against the list. Maybe it will help.

Rosenzweig, Roy, "Scarcity or abundance? Preserving the past in a digital era," *American Historical Review* (June 2003). (downloaded September 5 from chnm.gmu.edu/assets/historyessays/scarcityp.html)

This discussion of digital information and preservation begins with an oddity: the "Bert is evil" website disappeared from the web on October 11, 2001, for reasons spelled out in the essay. It's one of many examples of how readily digital resources can vanish. If you've looked at digital preservation strictly from a librarian's or archivist's perspective—or if you haven't thought about it at all—you may find a historian's perspective refreshing.

Noting that "Bert is evil" can still be found on the Internet Archive, Rosenzweig cautions that IA's existence "depends largely on the interest and energy of a single individual" and "has put the future of the past—traditionally seen as a public patrimony—in private hands."

There's a twofold problem, alluded to in the title. "Historians need to be thinking simultaneously about how to research, write, and teach in a world of unheard-of historical abundance and how to avoid a future of record scarcity." He notes that librarians, archivists, and computer scientists have been talking about these issues—but not historians, although "the split of archivists from historians is a relatively recent one."

There's a lot more here. He knows that some claimed losses never really occurred (almost all of the 1960 U.S. Census records have been recovered) and that software use longevity *may* be more of an issue than media life. That's a little tricky, though. He notes that Microsoft "only supports its software for about five years," but contemporary Microsoft applications will read and write files from at least eight years back, and readily available translation programs go much further back. Yes, it's an issue; no, it's not the same issue as software generation lifespans. (A later example, at a 20-year remove, is more convincing: That is, if a long-inactive writer died this year, leaving a pile of 5.25" diskettes containing manuscripts written in WordStar for CP/M or, say, Electric Pencil, would it be feasible to read and convert the files? Yes, it almost certainly would. Would it be easy or cheap? Unclear.)

The article covers a lot of ground. It includes some quotes I regard as silly, but you should draw your own conclusions. **Recommended**—and don't be too startled by the 21 pages of small print. The last nine pages are notes.

Spolsky, Joel, "The absolute minimum every software developer absolutely, positively must know about Unicode and character sets (no excuses!)." October 8, 2003: www.joelonsoftware.com

Disclaimer: I work for RLG. RLG is a founding member of the Unicode Consortium, and a colleague and long-time friend has been secretary of the Consortium for many years. RLG is the first (and, I believe, still the only) bibliographic agency to support the full set of nonroman scripts in common U.S. use for cataloging and display. We converted Eureka to Unicode display years ago—once the browsers started handling the heavy lifting (such as displaying Hebrew right to left).

I know people who helped make Unicode possible and understand it deeply. It's likely that those people "are going to find [Spolsky's] entire discussion a little bit oversimplified." He knows that: I just quoted the rest of a sentence that starts "Before I get started, I should warn you that if you are one of those rare people who knows about internationalization, you…" Or maybe they'd appreciate the informal and (I believe) essentially accurate commentary here, which I **recommend** if you want to know just a little bit about why Unicode matters.

"Did you ever get an email from your friends in Bulgaria with the subject line "???? ?????? ??? ???"?" That's one of the problems Unicode is designed to solve. Ever look at the start of a Web page and see something like this: <html><head><meta hettp-equiv="Content-Type" content="text/html; charset =utf-8">? That's a declaration that the page is in Unicode, encoded in UTF-8—which is a convenient way of transmitting good ol' ASCII (the first 128 characters in ANSI) as single bytes while still trans-

mitting the full richness of Unicode as needed (with other characters transmitted in two or more bytes each).

This is a good, reasonably brief, very informal discussion. It won't tell you how to use Unicode. It *will* suggest why Internet Explorer *sometimes* seems to handle non-English pages pretty well and sometimes screws them up entirely—a remarkable tidbit that I won't spoil for you. Go read it. Maybe you'll need to learn more…at which point you'll wind up referring to slightly less humorous articles by my knowledgeable colleagues and others.

A Library Stuff Perspective

Hysterical Librarians, Attorneys General and Section 215

Begin with Anita Ramasastry's September 2003 "Cyberlaw" essay at *FindLaw's Modern Practice*, "Why the ACLU is right to challenge the FBI's access to library, bookstore, and business records under the USA PATRIOT Act." That suit was filed on July 31 in Michigan federal court—the *first* suit to challenge provisions of Section 215, the portion of the act that authorizes FBI searches of records from businesses, libraries, and bookstores.

The ACLU considers Section 215 unconstitutional. Whatever Attorneys General (AG) may say, this section applies to everybody—visitors, permanent residents, U.S. citizens, whoever. Although the immediate plaintiffs (Arab-American advocacy and community groups) may be concerned about religious and ethnic targeting, the broader concern is the broad reach and lack of prior notice—and, for librarians at least, the fact that Section 215 overrides confidentiality provisions that most states have adopted. Here's Ramasastry's quick comment:

> In my view, the ACLU and the plaintiffs are entirely right to have brought the suit. Americans should not have to worry that the FBI is rifling through their personal belongings—learning what clubs we belong to, what charities we give to, and what books we read.

She notes that Senator Russ Feingold introduced the Library and Personal Records Privacy Act on the same day the suit was filed. (See *Cites & Insights* 3:12, p. 3, for a brief note on this act—which had "Booksellers" added after "Library" by the time I discussed it.)

The suit claims a violation of the Fourth Amendment because Sec. 215 greatly expands the FBI's power to obtain records and "other 'tangible things'" of people *not* suspected of criminal activity—without warrants, criminal subpoenas, notice, or showing of probable cause that a crime has been committed. All the FBI needs to do is certify that the records are related to "an investigation to protect against international terrorism or clandestine intelligence activity." *Not* that the person is suspected of being a terrorist—only that the records are somehow related. Are citizens and permanent residents protected if their activity is protected by the First Amendment? Not really: The Act only protects them if they're the *target* of the investigation. Ramasastry also notes that the lifetime gag order in Section 215 directly targets speech. This article brings out the point that the Justice Department has issued misleading statements—e.g., suggesting that Section 215 doesn't apply to citizens. That's not true.

The Proven Tactics in the Fight against Crime

That's the title of John Ashcroft's speech to the National Restaurant Association on September 15, part of the Great Flying Patriot Tour. Why would I be concerned with an association like this? Consider these four paragraphs, barely into the speech:

> If you were to listen to some in Washington, you might believe the hysteria behind this claim: "Your local library has been surrounded by the FBI." Agents are working round-the-clock. Like the X-Files, they are dressed in raincoats, dark suits, and sporting sunglasses. They stop patrons and librarians and interrogate everyone like Joe Friday. In a dull monotone they ask every person exiting the library, "Why were you at the library? What were you reading? Did you see anything suspicious?"
>
> According to these breathless reports and baseless hysteria, some have convinced the American Library Association that under the bipartisan Patriot Act, the FBI is not fighting terrorism. Instead, agents are checking how far you have gotten on the latest Tom Clancy novel.
>
> You may have thought that with all this hysteria and hyperbole, something had to be wrong. Do we at the Justice Department really care what you are reading? No.
>
> The law enforcement community has no interest in your reading habits. Tracking reading habits would betray our high regard for the First Amendment. And even if someone in the government wanted to do so, it would represent an impossible workload and a waste of government resources.

Ashcroft goes on to note that there are 11,000 FBI agents and, according to ALA, 117,400 libraries in the U.S.—visited more than one billion times last year, checking out about 1.7 billion "books." "The hysteria is ridiculous. Our job is not." Then, after claiming, "No one believes in our First Amendment civil liberties more than this administration," Ashcroft goes on to spend most of the speech talking about reductions in the overall crime rate before returning to terrorism. (He says the administration's tactics "prevented *any* terrorist attack over the past two years," which will surprise the heck out of our soldiers in Iraq and those rebuilding Iraq's infrastructure and, for that matter, U.S. businesses bombed by certain groups—but I guess he means "successful domestic attacks by known international terrorist groups.")

Let's go back to those four paragraphs. When have you ever heard any librarian or anyone else suggest that the FBI was "surrounding" local libraries or that agents would "stop patrons" or "interrogate everyone"? This breathless report comes from one source and one source only: John Ashcroft. It is as extreme a strawman as I've seen. Has ALA suggested that "the FBI is not fighting terrorism" or that "agents are checking how far you have gotten on the latest Tom Clancy novel"? No, any more than ALA was "convinced" by "some" to do so. Then, after the direct insults to librarians and ALA, we get the key statement: "The law enforcement community has no interest in your reading habits." Based on that sentence, Ashcroft should *support* a change to Section 215 that would exempt libraries.

Has the FBI ever evinced an interest in people's reading habits? You bet they have, back in the 1970s and 1980s (and before); that's the primary reason that every state enacted patron confidentiality laws. It happened at UC Berkeley when I worked there; it happened at dozens of other libraries.

Immediate Reaction

Curt Anderson of AP invited Emily Sheketoff of ALA's Washington office to respond to Ashcroft. She noted that library records *should* be treated differently than business records for privacy and free speech reasons—and, at a minimum, the Justice Department should *publicly* say how often FBI agents subpoena library records. "They are not taking this issue seriously, and the American people are upset about this. At least they could give us some idea of the breadth of the problem." Another Justice Department official claimed that the reason for FBI agents to want library records is "to track use of its publicly available computers." In which case, again, circulation records could be exempt—and it's a silly exercise. Many libraries don't have signup sheets for computer use—and those that do are unlikely to retain the handwritten sheets for all that long. It's possible that some libraries have exhaustive computerized records of who used public-access computers when—but it's not likely, even *before* this summer.

ALA president Carla Hayden issued a formal response the next day. In part, she said, "Because the Justice Department has refused our requests for information about how many libraries have been visited by law enforcement officers using these new powers, we have focused on what the law allows. The PATRIOT Act gives law enforcement unprecedented powers of surveillance—including easy access to library records with minimal judicial oversight." She notes that the act overrides "the historical protections of library reading records that exist in every state." Later, Hayden makes one of those crucial distinctions that the Justice Department seems to ignore. "Over the past two years, Americans have been told that only individuals directly involved in terrorism need be concerned. *This is not what the law says*. The act lowers the legal standard to "simple relevance" rather than the higher standard of "probable cause" required by the Fourth Amendment." [Emphasis added.]

> In March 2003, the Justice Department said that libraries had become a logical target of surveillance. Which assurance by Mark Corallo are we to believe?

Noting another "assurance" that the law only affects non-citizens, she repeats: "This is not what the law says." She goes on to recount FBI library contacts going back to the McCarthy era and offers concern that Ashcroft "should be so openly contemptuous of those who seek to defend our Constitution."

When the Scapegoat Strikes Back…

The next day, John Ashcroft called Carla Hayden. He "expressed his concern that people have misunderstood his commitment to civil liberties and committed to declassify the Justice Department report on Section 215."

So he did: "The number of times section 215 has been used to date is zero." Ashcroft disclosed this "to counter the troubling amount of public distortion and misinformation" about section 215."

The only distortion had been Ashcroft's claims about ALA and library attitudes: Neither ALA nor any librarian had offered a specific Section 215 number, since to do so would be a crime.

The chronology gets a little muddy here, but it would appear that Congressman Bernie Sanders released a press release after Ashcroft *promised* to make the disclosure, but before he actually did so. Sanders' quote:

> This is an important first step, but now Congress must work to amend section 215 of the Patriot Act. The issue is not just what the Department of Justice has done in the past but what it could do in the future as a result of this dangerous provision. The bottom line is that the Federal Government should not be able to walk into a library or bookstore without probable cause and obtain the reading records of the American people.

The next day, the info-commons.org commons-blog had a commentary (presumably by Frederick Emrich) on Ashcroft's "zero" including these notes:

> Let's be clear about a couple of points. The veracity of a Justice Department memo is open to debate. The appropriate vehicle for determining applications of the law with at least some reliability is Congressional oversight including open hearings. Furthermore, the number of times the law *has* been applied has little bearing on the fact that it *could* be applied. Civil rights carry no weight if the law allows the government to violate those rights, whether authorities actually choose to do so or not. Finally, if the government has never used these provisions, perhaps it does not need them at all. … [Quoting Emily Sheketoff:] "If this number is accurate, then they have demonstrated that there is no need to change the tradition of protecting library patrons' reading records."

ALA released another press release the next day. Carla Hayden offered surprise at learning that Section 215 had never been used, "particularly in light of previous statements from the Justice Department." In December 2002, assistant AG Daniel Bryant said information had been sought from libraries on a voluntary basis. (The March 2003 comment was noted above.) In May 2003, assistant AG Viet Dinh said federal agents had visited about 50 libraries. The release also renewed the call for Congress to restore the historic protection of library records.

Through the Wayback Machine

Here's a Golden Oldie, from *USA Today* for March 25, 2002: "FBI checks out library records of terrorist suspects." (Yes, it's still available on the web—www.usatoday.com/news/sept11/2002/06/25/fbi-libraries.htm) A University of Illinois survey of 1,020 public libraries found that 85 libraries had been asked for information about patrons related to Sept. 11—and in some cases directors reported that they were specifically instructed not to reveal any information about the request.

A Little Bit Later

Brent Hightower wrote an interesting piece in the New Orleans *Times-Picayune* on October 2, 2003: "Librarians, enemies of freedom? Now that's hysterical." Riffing off Ashcroft's "baseless hysteria" line, Hightower—a library school student and staff member at the paper's library—puts together a solid defense of librarians as *defenders* of freedom, including the freedom to read without government oversight. He also offers an interesting example of how FBI intrusion could go awry:

> Imagine a busy branch library on a Saturday afternoon. Mom's book group has tackled a novel about Muslim women, so for background Mom checks out books on culture, politics and religion in the Mideast. Then she helps her sixth-grader do a Web search for a report on Louisiana's chemical and shipping industries. … How do you think that Saturday's records would look to the FBI?

"Librarians seek to provide for the free flow of information on an equitable basis to all members of the community. These do not sound like enemies of freedom to me." A good piece.

And another, from Mike Argento in the October 9 York *Daily Record*: "Librarians keep quiet on terrorism, pornography." This was after *another* AG, Ed Meese, said on television that "Librarians are more interested in pushing pornography than fighting terrorism." (You *knew* pornography would get mixed in here somehow, didn't you?)

"Librarians? Pushing pornography? This was news to me. But I figured Ed Meese would know what he's talking about. After all, he was attorney general. He couldn't possibly be a complete idiot. You don't get to be attorney general by saying moronic stuff on national TV."

"OK. Let me rephrase that."

So Argento went to the library to do some investigative journalism—you know, see how the librarians were "pushing pornography" and whether they were "fighting terrorism." As to the latter:

> I saw librarians doing absolutely nothing to fight terrorism. They were doing librarian kinds of things, such as looking stuff up, shelving books, checking out stuff. None of them—none—were doing anything at all to fight terrorism. None were scouring the globe for Osama bin Laden. None were invading Iraq. None were even spying on American citizens to see whether they were reading subversive materials such as Al Franken's new book."

So, on this count, "Ed Meese was right." Then he checks out the pornography—and, as he noted, Ed Meese spent a *lot* of time checking out pornography when he was AG. He had the special task force investigating it—and achieving absolutely nothing. "Now, I've been to Martin Library, geez, must be thousands of times, and I've never seen the pornography section." But "unlike Ed Meese, I didn't know that libraries existed to push porno…. Maybe I just missed it."

So he asked Lora-Lynn Stevens, director of special projects and collections, to ask, "Where do they keep the pornography because I haven't been able to find it." Her response: "What?" Followed by, the library didn't have a porn collection, "last I checked." And Argento offers the only reasonable followup:

> Next thing you know, Ed Meese will be saying something like "Bartenders are more interested in selling liquor than fighting terrorism."

Which he'll have to check out.

Who Cares?

Why bother with all this a couple months later? You need to remember the chronology and the extent to which government officials feel free to scapegoat librarians. (As *American Libraries* reported in its November 2003 issue, Mark Corallo of the Justice Department said Ashcroft's speech was intended not as an attack on librarians but on "such groups as the American Civil Liberties Union"—and claimed ALA "has been somewhat duped by those who are ideologically opposed to the Patriot Act." You know, ideologues like people who believe in the Bill of Rights. So Justice decided it was OK to *admit* attacking ACLU, but librarians are just innocent dupes.)

- It wouldn't hurt to support moves to reform Section 215, particularly since the government claims it's never had to use the library portion. If the first two years after September 11 didn't require such uses, why would anything other than PermaWar and an Eternal Temporary Emergency Suspension of Civil Liberties require that the powers be there? You *know* Ashcroft wants the Patriot Act powers to be permanent—and wants more besides. Oh, I forget: He also assures us that no one believes in First Amendment liberties more than this administration. So it's OK.
- Assuming that you take Ashcroft at his word, is it reasonable to suspect that FBI agents could and did use the hidden-hammer approach: "You know our Section 215 powers? We need some information and we need it *now*. Do we really need a subpoena, or can we just get this out of the way?"

Feedback: Your Insights

First, two brief followups. "Getting that article: Good news," the lead essay last month, noted that the OAI harvester feature of one particular OpenURL resolver only offered journal and author searches against OAIster. A day or two after that issue appeared, the resolver—at the Université du Québec à Chicoutimi—was changed to add (or restore) a title-phrase search. It works beautifully. (Look for "The Crawford Files" in *American Libraries* for February 2004 for more comments on this OpenURL breakthrough.)

I should also run a partial correction to one of the assertions in the CIPA special: That no current censorware attempts to screen Web images based on image content. While that's true, it's also true that one company *claims* to block "explicit adult images" from email based on images rather than text. SurfWatch touts the "Virtual Image Agent" and makes the following claim:

> Using more than 22,000 different algorithms, the Virtual Image Agent can determine whether an image file attached to an e-mail contains explicit adult material. It can tell, for example, whether a flesh colored photo is an image of a person or something else; such as, sand or wood.

Since this sales pitch is for corporations looking to "save bandwidth" and avoid liability for "online voyeurism" by controlling email in the workplace, there are no apparent First Amendment issues and this doesn't relate to CIPA. Is the claim for the software plausible? I have an opinion, but since I'm not a grand master of algorithmic design, I won't offer it here. (Thanks to Will Stuivenga for passing on the information.)

Alex Bienkowski on Publishers and Preservation

> It seems to me that there has been a shift, quite a considerable one in fact, in where publishers stand on preservation. Many librarians are cautious about welcoming e-publication too enthusiastically because there seemed to be no real answer to the problem of preserving digital information. Publishers were rather off-handed about this: "We publish, we don't preserve." Preservation was felt to be a library problem and if librarians were worried they could buy the print, bind and store it as before.
>
> But now that major publishers are "retrodigitizing" their journal backfiles, it seems to me that they have quietly moved over into the preservationist camp, perhaps without even realizing it. It might be more accurate to say that they have assumed a community of interest with librarians that they didn't have before. If journal backfiles are business resources, and if selling access to them is a valuable revenue stream, these backfiles must be protected.
>
> I haven't seen any discussion of this shift, which I think is just as momentous as the Reversal of Alliances in the 1750's, which makes it all the stranger

to see it so little mentioned. If publishers decide to centralize and digitize and are not to see their assets decay and vanish, they will have to fund some serious R&D and there could be some beneficial fallout for libraries.

Then again: If publishers centralize and digitize, does this not affect their support for or acquiescence in projects such as LOCKSS? Why should they allow dispersal of their texts across the LOCKSS scheme, when they are trying to sell access to the backfiles kept on their servers? Your views?

Here's my initial response:

- The true long-term interests of a publisher will never be longer than the life of the publisher, where libraries have a longer-term view.
- In practice, I would impolitely question the true commitment of any commercial publisher to "preservation" for even one year past the point that sale of access to the backfile ceased to be profitable
- "Selling the backfile" continues to be a losing situation for libraries, since you don't actually *buy* the backfile, you buy some form of temporary access
- If publishers are willing to *sell* the backfile—that is, guarantee permanent access once an institution has paid the fee—they should have no objection to schemes such as LOCKSS, if such schemes have proper authorization routines built in (which I believe they do).
- These are my own opinions. I could be wrong. But at this point I remain to be convinced that it is plausible for a for-profit business to be regarded as a long-term guarantor of preservation: It conflicts too fundamentally with the rules of for-profit business.

Some for-profit businesses are in the long-term retention game, although even there "long term" may be the wrong name. My current understanding is that certain very large publishers who are actively "retrodigitizing" are working out (or have worked out) contracts to deposit full digital archives at institutions that *can* provide long-term preservation, such as national libraries. I'm sure I'll hear better-informed responses!

Peter Suber on Open Access

I'm going to summarize good points raised by Suber in an email conversation (flowing out of the November issue) rather than quoting at length. Suber notes, correctly, that open access literature doesn't have the "available copy" problem discussed in "Getting that article: good news," since open-access copies are available everywhere people have connectivity. We seem to agree that there continues to be a role for professional indexing and abstracting—and that open access could result in *more* indexing-and-abstracting services focusing on the freely available literature. Suber does note that work is needed to assure that open access articles are authenticated—so that, when you reach something via Google (or OAIster, for that matter), you can be reasonably certain that it is what it claims to be. And there's agreement that OpenURL compliance makes OAI compliance far more valuable.

Seth Finkelstein on a Variety of Topics

Here again, I'm going to excerpt and summarize Seth's set of clarifications on some of the points made in the November *Cites & Insights*.

- On "six degrees of separation": The *concept* that you can get from anyone to anyone in six transitions is a graph-theory mathematical result. "Given a graph of six billion nodes, and each node connected to (a few hundred? a thousand?) or so total other nodes, what's the average length of the smallest path between two nodes? I don't have a reference to the exact answer, but it's low." But knowing that it's *possible* to get from anyone to anyone else doesn't mean it will happen: That is, *if* a path completes (every person is being a router), it needs only a few hops. But don't expect many paths to complete. Overall, very few people may be interested in being routers. Thus, the low results of the study: Most people just didn't care.
- On compiling lists of child porn sites: "The key point I was trying to get across was that there's no legal paradox in censorware companies compiling lists of child porn sites… But authorities' disbelief can kill any Open Censorware blacklist project." My discussion was oversimplified.
- On the Google Dance: I misunderstood the nature of the "dance." Changes in Google's pagerank take place every 30 to 45 days; that's not the same thing as revisiting pages. My bad—I just didn't understand what Notess was getting at. So: Google visits cites.boisestate.edu ten times a day (13 times a day last month) but will not change the page rank more than once every 30 to 45 days.
- On EFF's "It's music to our ears" website: Seth Finkelstein reads the site as celebrating file-sharing as a technical achievement.

Jerry Kuntz (et al) on Federated Searching

Jerry Kuntz posted a provocative commentary on federated searching to Web4Lib on October 1, commenting on Roy Tennant's *Library Journal* col-

umns. I asked permission to use some of it, which Jerry happily provided. I intended to use it in the November *Cites & Insights*, but failed to do so—partly because of possible conflict of interest on my part, partly because I wasn't sure what I wanted to say about it, partly because of space. While I still suffer from some conflict of interest (as lead designer for Eureka), I think Jerry's comments are worth passing on—and since then, there have been two interesting additions to the "conversation." First, excerpts from Jerry Kuntz' posting:

> Roy Tennant's most recent article [on federated searching], although titled "The Right Solution," does cover several challenges to implementing federated searching that make you wonder if it's a solution at all:
>
> - The de-duping challenge
> - The relevance ranking challenge
> - The target configuration workload
>
> …and the article concludes with a caution based on the Searchlight project that, "We realized that an all-encompassing solution does not necessarily serve key-user needs well…the undergraduate…typically needs only 'a few good things' to cite… The graduate student or faculty member…wants much more thorough coverage. … Tailored portals [specific to a topic area or a select group of databases] may be a better solution."
>
> To the problems listed above I can add more reasons to hesitate over this technology:
>
> - Many of your electronic resources were not designed to use keyword searching as the primary navigation method, or in some cases may not use keyword searching at all. Which means you'll still have the problem of directing users to the appropriate resources both within the federated search and outside it.
> - A federated search tool can only support very basic search result sorting and limiting options, compared to the sorting and limiting options available in each resource's native interface.
> - There may be vast differences in the corpus word frequency among the indexed resources you're searching that will skew results. For instance, you'll find vast inequity in the frequency of keywords from a library catalog compared to an index of full-text articles from thousands of journals with a 20-year backfile…compared to a comprehensive web index like Google.
> - If you're using a federated search interface separate from your library catalog, but want to include your library catalog as a target, then you may be sacrificing features unique to the library catalog, like the ability to place holds or to have users view their own library record.
> - Different databases may be indexing different taxonomies that will skew results if users search on terms included in those taxonomies. They might get a ton of results from one resource, but none from another, because different subject terms have been applied in the indexing.
> - Authentication: Using a federated search assumes that all included resources are licensed for use by all potential users.
> - Z39.50 configuration differences. Yes, Z39.50 is a standard, but it's a standard that sure has left a lot of leeway for inconsistencies in the way one target will offer up results compared to another, especially when using boolean operators.
>
> I'm left wondering whether federated search products have been over-hyped and oversold, and whether the "holy grail" of one-stop searching is still many years away, and may need to incorporate an AI front-end that will interview the user before plopping a keyword search box in front of their nose.

My immediate comment dealt only with the suggested long-term solution (which actually harks back to early expert-system experiments in libraries). I believe there's strong anecdotal evidence (and some that can be inferred from studies) that each extra step in a search-and-retrieval process loses a significant portion of the searchers. That is, they just walk away. I certainly have local statistical evidence to suggest that *reducing* the number of steps in a search-and-retrieve process correlates with a substantial increase in the amount of effective searching, but it's hard to prove causation.

Jerry's third item on skewing results also applies to the Amazon Search In the Book. For searches that aren't specific in vocabulary, there's a tendency for irrelevant results to swamp real results.

In mid-October, Amanda Etches-Johnson offered her own commentary on "metasearching," another term for federated or distributed searching. Go to www.etches-johnson.com/bibliolatry/, look for October 10 in the archives. Briefly, while she's "not one of those librarians who holds on to the High Holy Principles of Librarianship at all costs," she's not impressed with the cross-database searching she's used even within a single vendor's databases. "Different databases use different syntax, indexes, and thesauri, so until vendors start making them all exactly the same, one size will just not fit all." Etches-Johnson is clear about not being hung up on teaching moments and the wonders of reference interviews: "My misgivings about metasearching aren't based on me not getting the warm & fuzzies on the reference desk, they're based on imperfections in the technology." (I would only add that perfecting the technology—somehow getting all databases to work identically, which is pretty much impossible—won't do a thing to solve swamping/skewing problems.)

Finally (for now), and most astonishingly (in my opinion), the October 2003 *Information Today* has a list of "the five most commonly repeated misconceptions about federated searching"—and it's from WebFeat, one of the biggest providers of federated searching technology. You can look up the whole writeup (www.infotoday.com/it/oct03/hane1.shtml), but here are the second, third, and fifth "misconception" and excerpts of WebFeat's comments:

> **2. De-dupe really works.** *Reality*: True de-duplication is virtually impossible. In order to de-dupe, the engine would have to download *all* search results and compare them. (The article goes on to say why that's pretty much impossible, and could take *hours* to yield a result if it was possible.) Vendors that claim to do true de-duping usually are just de-duping the first results returned by the search.
>
> **3. Relevancy rankings are totally relevant.** *Reality*: It's impossible to perform [such a relevancy ranking]. (The explanation has to do with the way searches are returned and the fact that full text and abstracts, which would make relevancy ranking more useful, aren't available to federated search engines.)
>
> **5. We don't make your search engine. We make your search engine better.** *Reality*: *You can't get better results with a federated search engine than you can with the native database search*. [Emphasis added.]

In each case, there's more—but it would be nice if people who think federated searching is a silver bullet would at least pay attention to the second and third items. In practice, bibliographic records aren't good candidates for typical relevance engines: There just aren't enough words in the records for the word-frequency analysis that passes for relevance. (Google and competitors use network analysis as well; there's no real equivalent for bibliographic retrieval.)

I'm sure this topic won't die.

Trends & Quick Takes

Where Are My Gigabytes?

The list of pathetic lawsuits seems to be as long as the list of outrageous patents. Here, from Reuters courtesy of *Wired News* on September 18, is one more: "(Hard drive) size does matter." Four Los Angeles residents are suing eight of the largest PC makers claiming, "their advertising deceptively overstates the true capacity of their hard drives." Naturally, they want class action status; although the brief story doesn't name the law firm, even non-cynical people might believe that it's a class-action specialist. (After all, how could four people plausibly sue *eight* computer companies for personal damages?)

The rest of you geeks out there already know the basis of the claim. Some computer measurements, specifically RAM, are stated based on powers of two. A kilobyte is two to the tenth power bytes, or 1024; a megabyte is a kilobyte squared (two to the twentieth power), or 1,048,576 bytes. A gigabyte is two to the thirtieth power: 1024 megabytes, or a little more than 1.073 billion bytes.

For quite some time now, hard drive capacity has been stated in billions of characters—unfortunately, abbreviated as GB. One of the footnotes in Gateway's catalogs and most of their advertising notes: "Hard drive accessible capacity varies: GB = 1 billion bytes." Similar phrases turn up in IBM, Sony, and HP ads.

"According to the lawsuit, computer hard drive capacities are described in promotional material in decimal notation, but the computer reads and writes data to the drives in a binary system. The result is that a hard drive described as being 20GB would actually have only 18.6 GB of readable capacity."

That first sentence is gibberish. I suppose you could sue the computer makers for not describing a 20GB hard drive as storing two to the 37th bits, more or less, since that would be the closest you could come in a "binary system." And I'm shocked that the law firm hasn't extended the suit to Microsoft, Apple's software division, and Linux makers, since file systems take away *even more* of the supposed capacity! Do the OS makers warn you that every thousand files you store in XP will, on average, *rob* you of a megabyte of storage that you *paid good money for?* And do Microsoft and Apple warn you that, between indexing (turned on by default in XP, and I suspect Apple OS does something similar) and the OS itself, there are probably billions of bytes of storage that *you paid for and can't use*? The horror!

The lawsuit wants an injunction against "unfair marketing practices," an order requiring defendants to disclose these nefarious practices (which at least five of the eight defendants already do in every ad I've seen), "restitution," disgorgement of ill-gotten profits and attorney's fees. Attorney's fees: That is, I suspect, what this suit is all about. Otherwise—well, heck, give the plaintiffs ten bucks each (the worth of the "missing" 10GB space on a very large drive) and tell them to go away. Ill-gotten profits? Since every PC maker uses the same standards, and since there's very little profit in the commodity parts of a PC such as hard disks, I don't see it.

I could *almost* see this suit if the only defendants were Dell and other companies that have stopped using the footnote in their ads. I hope the suit will be laughed out of court, but I suspect there will be a

settlement that's lucrative for the law firm and meaningless for anyone else. That's a shame.

The Segway Saga Continues

"[The Segway] will be to the car what the car was to the horse and buggy." That's Segway's inventor Dean Kamen in a December 10, 2001 *Time* interview—and it's the lead in Richard Roeper's Chicago *Sun-Times* column for September 30, 2003. He follows with a CBSNews.com note on the recall of Segways "because riders have been injured falling off when its batteries are low." Then: "Guess we might want to hold off on that whole Segway-car-horse-and-buggy comparison for a while."

As Roeper (*that* Roeper, if you're into movie review shows) notes, the most embarrassing thing for Segway's maker is not that three people have been injured on Segways, but that no more than 6,000 of the scooters—oh, sorry, personal transportation devices—have been sold in 21 months of production. Take away fleet sales (Atlanta's hospitality folks, US Postal Service test run, etc.) and the number would be even smaller. As Roeper says, "How many Segways do you see during your typical day?"

Kamen promised sales of 40,000 Segways *a month* back in 2001. It's possible that sales in the second year of production are on the order of 400 a month: That's two orders of magnitude difference.

Roeper's piece discusses Dean Kamen's genius (he invented the first portable insulin pump and portable dialysis machine) but also his goofiness. "We're talking about a man who divides his time between a hexagonally shaped, 30,000-square-foot house in New Hampshire and his own island off the coast of Connecticut, which has its own money, flag and navy."

Roeper was apparently as skeptical as I was when the Segway was first touted as revolutionary. "I just couldn't imagine how a 70-pound, $5,000 glorified electric scooter was going to change the world." He notes all the good press, including the gee-whiz stories from journalists who got to ride one. "Then again, how many of those reporters actually bought one for their own use?" He doesn't buy the idea that Segways belong on the sidewalk, particularly in busy cities, but doesn't worry about it too much. In addition to being useless for most older people, cold-weather cities, poor people, and people who actually need to *go* somewhere (or are sensible enough to walk shorter distances), the Segway has one killer flaw for non-geeks: "When one rides a Segway, one looks like a major dork. By its very design, the Segway is uncommonly nerdy." And, you know, he's right: I have yet to see a picture of a Segway being ridden, or the real-life examples in Atlanta during ALA, where the rider didn't look too geeky even for my tastes.

The Blogging Iceberg

That title comes from a survey report by Perseus Development Corp., which claims that 66% of "4.12 million hosted weblogs" had not been updated in two months—and more than 25% of weblogs had no postings after the first day. The commentary, available at www.perseus.com/blogsurvey/, offers demographic details and other points—e.g., "active blogs were updated on average every 14 days" and just over 1% of the blogs surveyed were updated daily.

The most interesting part of this survey is the conclusion, "Nanoaudiences are the logical outcome of continued growth in blogs." The mathematics are tough to challenge, but they state an "average" in a field where averages are wholly meaningless. If we get to the point where 100 million people regularly read blogs, and each of them reads 50 blogs (an awfully high number), *and* if there are 20 million active webloggers, then the average audience per weblog will be 250 people. So?

The title? "An iceberg is constantly dissolving into sea water, and the majority of blogs started are dissolving into static, abandoned web pages." And this charming item, after noting that 90% of blogs come from people aged 13 to 29 (51.5% 13-19, 39.6% 20-29): "Blogging is many things, yet the typical blog is written by a teenage girl who uses it twice a month to update her friends and classmates on happenings in her life… Underneath the iceberg, blogging is a social phenomenon: persistent messaging for young adults."

The story includes a number of caveats. I'd add one more: Since 3,634 weblogs were actually studied, the results don't necessarily mean anything. Projections of sampling error only work when what's being sampled is homogeneous enough to follow plausible curves. I think the results themselves imply that this isn't true for weblogs. Further, some figures are stated with implausible precision: "106,579 of the hosted blogs were updated on average at least once a week." That's ridiculous precision, given that the sample size was less than one-tenth of one percent of the population. "Around 100,000 of the blogs" might be a reasonable statement, but the "579" is bogus.

More to the point, if weblog readership follows a power law, then any statements about weblogs as a whole are pretty much meaningless. I'd be inclined to believe that most weblogs with more than 500 readers—where the authors have enough feedback or

metrics to know they have that many readers—are updated regularly, and that most weblogs with fewer than a dozen readers cease fairly rapidly.

Seth Finkelstein commented on the Perseus survey after grumping about the hype and blather of BloggerCon. He doesn't think weblogs are likely to "revolutionize politics, overthrow journalism as we know it, or change the world into cyber-utopia." Finkelstein states the key reasoning error behind some weblog hype as follows: "People assume production is the same as audience." It's not, of course. "Everyone can't have an audience of millions. That's a simple mathematical fact." He goes on to ask a key question I've wondered about whenever anyone talks about weblogs—or any other internet innovation—overthrowing traditional media. "Are the media [that do] have an audience of millions going to just go away? Why would that happen?" Why indeed? New magazines are doing fine these days, and older magazines are recovering ad sales and in many cases building readership.

Amazonia Gone Wild

Fair warning: This particular topic—and I suspect most readers know what it is before I go into it—pushes a lot of my buttons, some of them unfairly. I'm no great fan of Amazon: I believe it harms local booksellers and I suspect its collaborative filtering is biased toward what they want to sell, and I *know* they manipulated submissions and information relating to one of my books. I'm no great fan of library doomcryers. And, of course, I'm no great fan of the concept that public libraries should be Just Like Amazon (or Just Like NPR, or providing most services only to those who wish to pay for them, or…). And my own quick tests of the feature in question yielded bizarre results.

I'm still not ready to attack the concept.

I *am* ready to poke fun at "The great library of Amazonia," Gary Wolf's lovefest article (posted October 23 at *Wired News*, and part of *Wired* for December 2003). But, as with most of what used to be "Cheap shots & commentary," I wonder if it's worth the space. How about this, in the third paragraph, just after extolling the wonders of the Internet Archive: "Books take time to transport. Their text vanishes and their pages yellow in a rash of foxing." Apparently, Wolf believes that digital resources such as the Internet Archive are *more* durable and permanent than books. With wisdom and knowledge on that scale, it's hard to know where to go. Of course Wolf is instantly captivated by the trial version of Amazon's book-search feature: How could it be otherwise. I'm impressed by Amazon's answer to copyright questions: "The company simply denies it has built an electronic library at all." A little later, we learn that Search Inside the Book will "undoubtedly fuel enthusiasm for overturning the current publishing and copyright regime."

After that, we get the usual KTD stuff—from Brewster Kahle, who really should know better than to say, "For most students today, if something is not on the Net, it doesn't exist." Wolf goes on: "Students, among others, are blind to the most important artifacts of human knowledge"—because they *don't know* there are such things as books. Right. And, further, the copyright issue is one of "political will"—which, I suppose, is true: With the right politics, rights of authors can simply be overturned.

We also learn that the current publishing system means "a title becomes inefficient at thousands of sales per year"—which may be true for the largest publishing houses, but is absolute nonsense for everyone else. Wolf claims Kahle's Internet Bookmobile "produces decent-quality paperbacks…for about $1 each," which makes me wonder just where he gets sheet paper and toner at such remarkably low prices (if the "paperbacks" are more than 50 pages long, at least). There's more, but this mostly reminded me why I rarely cite *Wired* articles.

The service itself had the cooperation of publishers but not authors. The Authors Guild looked at typical book contracts and concluded that publishers didn't have the right to consent without author permission. The email that the guild sent to its members (find it at www.authorsguild.org) is even-handed: "Whether your works should be in the program is hard to say." They point out that some books will benefit from the exposure—while for others, as Amazon originally implemented it, being able to print out up to a fifth of the book could gut sales. (As far as I know, Amazon has now changed the service so that it's either impossible or difficult to print out pages. That's the gist of a November 3 AP story, at least. I'm not willing to give Amazon the personal information needed to actually see pages.)

A story at LawMeme on October 30, while finding the feature interesting, points out that—because only registered users can actually see pages, and given Amazon's limitations on what you can do—Amazon is almost certainly tracking all usage of the feature. "This is an exercise in data-gathering that makes their old experiments with differential pricing look like child's play." The piece also points out that it's a true trade: You give up a whole bunch of browsing privacy in return for a new kind of information access. And ends: "All the same—and despite thinking that this is an insanely cool search applica-

tion—I still haven't gotten around to looking at any actual pages…"

Then there's Web4Lib, with a discussion that started on or around October 25, yielded a *bunch* of messages for several days, and tapered off by the end of the month. A variety of sensible perspectives were offered, along with a few that I find exaggerated. One person was impressed by the accomplishment of building the database but didn't think the feature was ready for prime time—and noted that it tended to screw up searches for book titles. Another, with an *extremely* specific query, loved the feature—and it's hard to argue with that enthusiasm, as long as the query is specific enough (and you don't mind that it covers a relatively small universe). One person was nervous about searching an unknown corpus—and another saw the sky falling on public libraries. Some people used the service and hated it (mostly because in-text results swamped book titles). That's just a sampling of comments, most of them from people who are a lot more thoughtful and knowledgeable on this topic than I am.

I don't know. I honestly don't know. Except this: I'm 99.99% sure that this service will no more doom public libraries than did Amazon itself.

Quicker Takes

- A sad story appeared in the *Daily Californian* on September 8. (That's UC Berkeley's student newspaper.) "Threatening the campus library's top-notch reputation, soaring subscription costs are chipping away at the library's journal offerings—a problem severely compounded by millions in state budget cuts." More than 600 journals were cut "or reduced" this year. The piece notes one extreme case of print pricing, *Chemical Abstracts* at $22,000 per year, and quotes a professor claiming that Berkeley no longer has access to journals "that many other far-smaller colleges have access to." This professor says, "It's really pretty bad." It probably is. It isn't helped by the lumping of libraries in with administration—so that a $36.5 million systemwide permanent cut affects the heart of the academic enterprise as well as administrative offices. Not that there's anything new about Berkeley cutting serials. In the early 1970s, I participated in a 10% dollar cut in serials expenditures. Back then, it was mostly cutting fat (reducing the *number* of subscriptions to *Biological Abstracts*, for example); these days, the cuts are far more serious. I do appreciate the closing quote from neurobiology professor Jeff Winer: "The word library and the word university belong together…it's a shame, it hurts research here."
- Remember Nancy Pearl, she of the (second) librarian action figure and the new book *Book Lust*? She's my age, and she offered a wonderful rule for dealing with books that don't live up to expectations, as recounted in a New York *Times* interview: "Nobody should ever have to finish a book they're not thoroughly enjoying, but you need to give the book a chance." Her original "rule of 50" was that you should read the first 50 pages before deciding to abandon the book. But as she got older herself, she realized "time was short, and the world of books is larger than ever." So here's the Amended Rule of 50: "If you're over 50, you subtract your age from 100, and that number is the number of pages you have to read." What a wonderful rule! Now I can give up after 42 pages if it just isn't happening for me…

The Library Stuff

Carlson, Scott, "Libraries' consortium conundrum," *Chronicle of Higher Education* (October 10, 2003).

How many consortia does your library belong to? Does the library play one consortium off against another? This article, **recommended** if you have easy access to the *Chronicle*, offers some startling figures and brings some of the complex world of library consortia to the broader academic audience.

The initial anecdote astonished me. St. Ambrose University's library paid $26,000 for two full-text databases in 1999 "when times were flush." Now, with a leaner budget, the library gets 16 full-text databases for $1,200 a year thanks to consortia (and grants secured by consortia).

Dillon, Dennis, "The librarian's world and welcome to it," *Chronicle of Higher Education* (September 19, 2003).

If you haven't already read this charming two-day chronology (in the *Review* section), go find a copy and read it. It's mostly about coping with faculty expectations for access or print subscriptions to obscure academic journals in a time of curtailed resources—but it's funny, lively, and makes the point much better through humor than it's usually made through argumentation. **Strongly recommended**.

Dunneback, Katie, "Weblogs as libraries," The Young Librarian (October 7, 2003). www.geo-

www.geocities.com/young_librarian/bloglibessay.html

We need young librarians who offer new perspectives. Is Katie Dunneback (the young librarian) seriously positing that weblogs are libraries? Well, here's the start of the essay:

> Conceptualize this: a weblog is a library. Yes, I hear you screaming. If all weblogs are libraries, then does that mean that all webloggers are librarians? Not really, especially since not all library workers are librarians.

Is this a "lamebrain idea" (as she calls it in the next sentence)? Yes and no. Sure, it's a little absurd to grant weblogs the stature of libraries—but *thinking* of weblogs in library terms is interesting. Dunneback does a good job for an informal blog essay. Am I recommending this essay? I found it thought-provoking; you might also.

Goans, Doug, and Teri M. Vogel, "Building a home for library news with a BLOG," *Computers In Libraries* 23:10 (November/December 2003). Downloaded from www.infotoday.com/cilmag/ November 3, 2003.

Most of the relatively few library-based weblogs I've heard of are in public libraries—and, as the authors note, "few libraries are actually [using weblogs]." Georgia State University's library needed a more dynamic web presence, and the authors decided that weblog technology might be the answer. They surveyed the field and some of the literature, thought about features, considered various hosting and software possibilities—and decided to roll their own, building the whole thing in house (using MySQL to store content).

This isn't one of those "let's just try it" situations. They planned it out, built two weblogs (one for general library news, once for science news), went through prototype designs, and went live in January 2003. Then they started considering additional features, marketing, branding, and all that.

It's an interesting story. As they say, even though they chose to build their own software, "setting up the technology may be the easy part. Blogs, like most Web text, will fail to attract repeat visitors if you don't keep the content current."

King, Donald W., Carol Tenopir, Carol Hansen Montgomery, and Sarah E. Aerni, "Patterns of journal use by faculty at three diverse universities," *D-Lib Magazine* 9:10 (October 2003). www.dlib.org

Here's another in the long series of King/Tenopir studies of reading patterns. This one compares journal use at the University of Tennessee while it was in a "transitional phase" from print journals to online full text, Pittsburgh with heavy full text but also extensive print, and Drexel with its primarily online-journal philosophy.

As with most such studies, it's important to read the study details before considering the conclusions. Two of the three surveys reported here involve fewer than 100 respondents, and all three involve self-reporting on the number of articles read over the last 30 days. Is it possible that the 18% to 50% of faculty (and staff) who responded might read more heavily than those who failed to respond, and that some people might pad their responses? Possibly, although the projected averages of 175 articles read (or skimmed, or glanced at) over a year by nonscientists and 216 by scientists don't strike me as high.

Even given the caveats that attend to surveys of this sort, the results are interesting—or, in some cases, not surprising at all. For example, almost all personal subscriptions are for print journals, even when electronic versions are available. Drexel has indoctrinated its faculty in the wonderfulness of e-journals and made them aware that the library funds access to those journals. And look at the numbers: *none* of the "other" sources for Drexel were "free Web journals," as compared to 5% to 9% at the other campuses. It's likely that most of those "free" journals at Tennessee and Pitt were actually library-licensed fee journals—and it's possible that some open-access journals were read by Drexel faculty who assumed them to be library acquisitions!

When it comes to readings from the library collection, the print-and-electronic percentages for Tennessee (78% print) and Pitt (75% electronic) seem sensible. The fact that 17% of Drexel reading was from the *print* journal collection is interesting, and makes me wonder how well Drexel's single-mindedness serves its faculty. (Consider also that Drexel's faculty read more from personal subscriptions than the other two: Is there a correlation?) These numbers do *not* imply that scholars are actually reading articles online. Given that Drexel faculty managed to get 57% of their articles from print sources overall, it seems likely that most articles read from electronic sources were also printed out.

Lancto, Craig, "Banned books: How schools restrict the reading of young people," *The World & I* (September 2003). www.worldandi.com/newhome/public/2003/september/mt2pubprint.asp

I wasn't sure whether this belonged in Good Stuff or Library Stuff, but it's *interesting* stuff. I was not aware that Ballantine Books "had been censuring and expurgating *Fahrenheit 451* for years" with-

out Ray Bradbury's knowledge or permission. I'm not surprised that, when the unexpurgated version appeared (in 1992), school officials in Irvine "immediately ordered teachers to black out 'obscenities' such as hell and damn before distributing the books to their students." And, as is frequently the case when censors claim to have overwhelming public support, it wasn't so: "Angry parents…forced the school to rescind its censorship."

This wide-ranging informal essay questions why some books are banned and challenged, the extent to which fear of censorship limits what gets published, and the extent to which "evil" in books is in the mind of the beholder. (Lancto points out that J.K. Rowling is a newcomer when it comes to mentioning witchcraft in books that children might read: "The earliest example of a book containing witchcraft that I found was the Bible. Although it has been widely banned for a host of reasons, witchcraft has not been among them." Well, there's also some highly erotic poetry in that book…

Recommended as a good read and non-librarian's view of an ongoing library issue. Lancto closes: "When critics insist that everyone else conform to their beliefs and opinions, when they stifle discussion about the perils that loom large on their children's horizons, they handicap the children and deny them the opportunity to understand the danger as well as the attractions of forbidden fruits."

Mann, Thomas, "Why LC subject headings are more important than ever," *American Libraries* 34:9 (October 2003): 52-4.

Who needs subject headings when you can use keywords? Only researchers who want to make sense of large collections and complex subjects—which, in many libraries, means almost everybody. Mann spells this out in detail and presses for subject browse displays as a vital part of any online catalog. As always with Thomas Mann, his views may be traditional but his reasoning is impressive and his writing convincing (at least to me). **Well worth reading**.

Admitted personal interest: Partly at my urging, Eureka has *always* channeled subject searches to subject browse lists—and, when you click on a subject heading in a record, you go back to that browse, *not* to a direct subject search. I believe Eureka was the first online catalog-equivalent to do the latter, and substantial evidence shows that both uses of browse benefit users.

Marcum, Deanna B., and Gerald George, "Who uses what?" *D-Lib Magazine* 9:10 (October 2003). www.dlib.org

This report on a national survey of information users in colleges and universities seems to have an agenda, referring to traditional research as "quaintly inefficient." After all, what modern student or faculty member would deal with the printed works in library stacks when they could "sit in the comfort of a home or office to do research"? Given the parties involved—the Digital Library Federation and Council on Library and Information Resources commissioned a study done by Outsell—it doesn't take a cynic to assume that the results will be interpreted to favor digital resources whenever that's plausible.

But I haven't looked at the study as a whole. Given that it's 893 pages long and includes 659 tables, I'm not likely to. According to the summary, that massive analysis is based on 3,234 half-hour interviews between November 14, 2001 and February 2, 2002. Even the summary stresses the caveat that everything's based on "what respondents *said* they do when searching for and using information" [emphasis in the original] and that some questions didn't yield enough responses to be very meaningful.

Briefly, most faculty and students are comfortable retrieving and using digital information—and how could it be otherwise at this point? It's equally unsurprising that almost all respondents use electronic resources for research some of the time.

Still, two-thirds of the faculty and grad students use print *resources* for research all or most of the time (73% for teaching)—and, although Kids These Days all demand digital everything (as we all know), even 52% of *undergrads* use print resources for coursework all or most of the time. (The figure was 72% for grad students.) More than 90% of them agreed that print books and journals "will continue to be important sources for me for the next five years."

Here's an odd one: "Nearly one-fifth reported using e-books." Was there a follow-up question, "What do you mean by e-books"? Not odd at all: More students use online indexes than printed indexes. Still, 59% of undergrads still use print abstracts and indexes, 93% use print books, and 81% use print journals (97% of grad students use print journals).

"Although higher education's users of information may frequently consult electronic resources, they are likely to print out what they find." "Likely" as in 77%--and only 28% said they "find reading information on screen satisfactory and rarely print out information." (Apparently, 5% are "likely" to print out resources but "rarely" do so.)

More figures? 86% of students feel that "my campus library meets most of my information needs"—a showing that some doom-and-gloomers should pay attention to. 55% still regard browsing

the stacks and journal shelves as an important way to get information—and only 35% use the library significantly less than they did two years ago. I say "only" because prevailing reports seem to be that Kids These Days never enter campus libraries at all. "Only" 16% said, "the Internet has not changed the way I use the library." I put "only" in quotes because that's the word the article uses—and I'm surprised the figure is that high, even for humanities people.

I do not understand why certain people insist on calling electronic resources provided by libraries "virtual," as this piece does. Libraries pay hundreds of millions of dollars for access to those resources. That's real money for real resources on real server farms, mostly winding up printed out on real paper. And that real money is a real problem, particularly for the continuing health of those print journal runs and books that 55% of today's students *still* find useful to browse.

What do students need more of from their libraries? Of reported responses (other than "nothing"), it's worth noting that 14% want more *print* journals as compared to 11% who want more e-journals—and 8% want more books. Are there lessons here for the digital library mavens? Apparently not, at least for these authors. They wind up by asking "how locally focused does it make sense for any content provider to remain?" and what it will take to "[dissolve] access boundaries to provide the professors and students of the future with the libraries of the world?"

The summary is **worth reading**. I'd love to see an independent analysis of the entire report, perhaps from someone who is not a committed digital library adherent. I won't hold my breath.

Nicholson, Scott, "Avoiding the great data-wipe of ought-three," *American Libraries* 34:9 (October 2003): 36.

Scott Nicholson is hot on "bibliomining," the library version of datamining. This "on my mind" piece urges libraries *not* to discard circulation data once items are returned. Instead, learn from the corporate world and maintain a data warehouse.

If libraries have the resources for data warehouses and datamining and if the process of cleaning transactional data replaces user identification with *non-unique* codes (demographic information, maybe—but with considerable caution about identifiability), then I don't have any argument with Nicholson's thesis. But the corporate world is hardly the exemplar of high regard for customer confidentiality! Data warehouse solutions that come from the corporate world would need careful checking to assure that it is truly impossible to resurrect borrowing histories. I wouldn't assume that to be the case.

I do wonder, however, in "times of tight budgets and corporate competition," to use Nicholson's phrase, just how many libraries are positioned to invest in data warehouses and datamining? (I also wonder what the "corporate competition" for a public library is, but that's another issue altogether.)

OCLC, "Libraries: How they stack up." 8 pp., downloaded from somewhere at www.oclc.org.

This graphic "report" has some interesting comparisons; if you haven't already seen it, take a look. For example, U.S. library expenditures are roughly comparable to the amount spent in bars and taverns, on athletic footwear, or on videos. Good public libraries return several times their expenditures in benefits—a 10:1 ratio at Phoenix Public, if you believe one methodology. U.S. libraries circulate about four times as many items each day as Amazon handles (and Amazon's far more than a book peddler these days), or about the same number of items as FedEx ships—and Americans visit libraries more than five times as often as they attend sporting events. Some of this is just "fun stuff," some of it may prove useful in discussing the value of libraries. **Worthwhile**.

Resnik, Wendy, "The 'L' word versus the 'I' word," *BiblioTech*. Downloaded from www.sir.arizona.edu/lso/ November 3, 2003.

I believe there's more than one *BiblioTech*; this one is a new "webzine by and for SIRLS students." SIRLS is the School of Information Resources & Library Science at the University of Arizona. So this is a library school student perspective—one that's both interesting and mildly provocative.

As a true believer in "L," I guess I should plan to attend Clarion University: it's the only ALA-accredited program in the U.S. that doesn't include "information" in the school title—while "twenty-eight percent" (in other words, 14) *omit* "library" or "librarianship." That's a shame, frankly; while Information Science may seem to be more prestigious, I still regard it as a somewhat vacuous field. Resnik does not—and I respect her argumentation.

She's almost certainly right that "universities…look down upon library science as a discipline lacking academic depth," although I suspect universities would overlook that failing if library school faculty could bring in big research grants. I wonder about her assertion that library science is about skills, while information science is about theory. I wonder even more about these two sentences:

> The science of information is a compelling and dynamic field not limited to librarians. Webmasters, programmers, information brokers, and, yes, librarians can all be taught under the same theoretical umbrellas.

Just at a guess, damn few webmasters or programmers are ever going to get MIS degrees—although I could be wrong. Information brokers? I'm not sure I know what they are.

I could pick more nits. "The nature of information is changing and the change is driven largely by technology." I'm unconvinced that the nature—the essence—of "information" is changing, although the techniques for storing and manipulating it certainly are. I'm less ready than Resnik is to buy the idea that "how to catalog books" is a skill that should be learned on the job—that there's no theoretical basis to classification and cataloging. But then, I don't have the degree, so what do I know?

While Resnik's reading for this commentary is narrow—surely *Library Journal* is not the only place where these issues are discussed—her presentation is clear and **worth considering**.

Webster, Peter, "Implications of expanded library electronic reference collections," *Online* 27:5 (September/October 2003): 24-7.

Webster brings an interesting combination of perspectives to this discussion: He's systems librarian at the Patrick Power Library of Saint Mary's University in Nova Scotia, but he also does reference work regularly as well as bibliographic instruction, or at least database instruction. He sees the advantages of electronic reference works but he's aware of the unintended consequences. For example, some Canadian publishers are discontinuing paper versions of reference works and he expects to see more of that: Cancelled print subscriptions make it impossible to continue print. For some works, that may be fine; for others, it's a problem.

Here's a slightly appalling commentary, and I don't doubt Webster's reporting:

> Not so long ago, a senior professor told me that he would no longer refer his students to the paper indexing and abstracting tool in his discipline. He knew that this was *unquestionably the best source* for the materials his students needed. However, he admitted that he had always hated the product and struggled with its multiple volumes and multiple indexes, even as a graduate student. Knowing that there were simpler online and full-text tools available, he did not feel he could force his students to use the paper abstracts, even though he knew the results would be significantly better if they did. [Emphasis added.]

I'm an optimist: I'll hope the professor at least *mentions* that the paper abstracts may yield better results for students willing to work harder. But I'm also a realist: Many students will never learn of the better resource unless they ask a librarian.

Webster notes areas—even in reference—in which the "superiority of books for many purposes is being reconfirmed," such as dictionaries for quick lookups and the *Europe World Yearbook*. He goes on to discuss the dangers in moving entirely to e-journals (including the loss of long-term ownership) and some cost and pricing issues for online resources. There's more; this is **well worth reading**.

West, Jessamyn, "The librarian is in and online," *Computers in Libraries* 23:9 (October 2003).

In this lively article, the Jessamyn West of librarian.net discusses her brief "career" as a participant in Google Answers and some lessons she draws for librarians. She sees three big differences between Google Answers and reference librarians: Google Answers researchers generally aren't librarians (and rarely pretend to be); question-answering isn't the same as reference work; and there are always people who prefer to pay for convenience. She discusses each of those three areas and adds a number of specific lessons for librarians.

Recommended, even if some of the lessons are ones you already know. It doesn't hurt to be reminded by real-world experience going outside the library—and maybe your unease about being usurped will be calmed. I love this sentence: "I know we librarians say that there are no stupid questions, but I believe there might be stupid question-askers."

The Details

Cites & Insights: Crawford at Large, Volume 3, Number 14, Whole Issue 43, ISSN 1534-0937, is written and produced at least monthly by Walt Crawford, a senior analyst at RLG. Opinions herein do not reflect those of RLG. Comments should be sent to wcc@notes.rlg.org. *Cites & Insights: Crawford at Large* is copyright © 2003 by Walt Crawford: Some rights reserved.

Original material in this work is licensed under the Creative Commons Attribution-NonCommercial License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc/1.0 or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

URL: cites.boisestate.edu/civ3i14.pdf