Cites & Insights

Crawford at Large

Libraries • Policy • Technology • Media

Bibs & Blather

Moving My Website

http://waltcrawford.name. If you're a traditionalist, waltcrawford.com will work. That's where my personal website is—and the old site (which now consists entirely of a note to use the new site) will disappear in a few weeks. (Right now, I'm paying $19.95 a month for a dialup account *purely* to keep that site alive, and that's just silly.) If you're wondering, waltcrawford.name is another part of the LISHost empire (actually, Jenny, there are *lots* of things we agree on, including hosting facilities), as is *Walt at Random* (but in that case, it's obvious from the URL).

To my own surprise, I migrated all the archived content from the old site to the new, including a couple of old essays I might wish to rethink or ignore. Those essays exist; I see no reason to pretend they don't. (For that matter, I wouldn't disown any of the essays; I just might write a little less vehemently in one or two cases.) What didn't migrate is the early *Cites & Insights* and the single more recent issue that was temporarily unpublishable at cites.boisestate.edu. I'm considering a *C&I* mirror at waltcrawford.name, but haven't decided anything.

Mostly, the point here is that—even though the first few entries under "Walt Crawford" on the major web search engines will probably still guide you there—my old website will disappear in the near future. Go to waltcrawford.name instead.

Where Have All the Readers Gone?

[Apologies, Peter Paul & Mary fans. Old folkies never die, we just fade away…] I'm writing this on August 1. The metrics for *C&I* since January 1 just appeared. It does look as though my excuse, er, reason for publishing a combined July/August issue was a good one.

Which is to say that readership for the first 18 days since that issue appeared was pathetic—enough to keep doing *C&I*, but about half what I'd expect in the first 18-20 days, and about a third of a typical issue's first-year readership. It's not that everyone was reading the HTML pieces, or at least not *those* HTML pieces: Only one piece in the combined issue was among the top 45 in year-to-date HTML readership, with fewer than 100 unique readers.

I'm not surprised. I'm reading less professional material than usual during the summer, and "serious" blog entries seem to have declined substantially. Many of you have the good sense to take unplugged summer vacations; others are too busy relaxing to spend precious time reading long essays on copyright balance or the perils of futurism. Maybe in the fall?

Inside This Issue

Clarifying My Stance on Copyright

I give up. Any reader of *C&I* who is fully literate and capable of understanding long sentences *knows* I'm a firm believer in creativity and in the rights of creative artists to earn money from their creations (if anyone wants to pay for them, to be sure). They should also know that I place American copyright in a Constitutional framework, which automatically means "intellectual property" is something other than an eternal, unlimited, completely unrestricted property right. (There are very few eternal, unlimited, completely unrestricted property rights, as it happens.)

I also believe current copyright law and practice is seriously overbalanced toward the interests of copyright holders—not necessarily the creators, but Big Media and others who control rights. I believe the

Life+70 years (or 95 years for work done for hire) term is absurdly long. I believe the lack of registration—in an era where online registration could be made free and easy—poses its own set of problems, making it difficult for publishers and creators to reach copyright holders in order to *pay* license fees. I believe expansion of copyright to define "derivative works" very broadly further damages creativity.

I've gone to some pains to clarify my stance in recent *Walt at Random* posts and *C&I* essays. As I was preparing the ©2 PERSPECTIVE in this issue, I found myself writing such strained comments as this:

> For example (drawing from a March 10 summary at FreeCulture.org, *which does not imply that I support FreeCulture's stance*):

No more. I've deleted extraneous "just because I quote someone doesn't mean I agree with them" notes and will continue to do so.

It's now clear that no amount of clarification will prevent deliberate misunderstanding and selective quotation on the part of those who wish to view me as an anti-copyright advocate. To the best of my knowledge, there's only one current case of someone deliberately misrepresenting my views. I'm no longer willing to waste my time and *C&I* space trying to prevent one jackass from braying.

A Light and Fluffy Summer Issue

That's what I had in mind here, knowing it would appear while people are still in that summer haze. Six or eight short sections; nothing that would require careful reading or make you think.

But *Cites & Insights* wouldn't be what it is if I planned it. You may find that this issue doesn't make you think or require careful reading, but it's certainly not full of brief essays. Quite the opposite: it's chunky.

Chunky issues are *never* planned. They just happen. In this case, the "biblioblogosphere" investigation turned out to be more interesting and complicated than I thought, and that's reflected in the length of the results (it could be worse: I could be including the spreadsheets within *C&I*, which would chew up several more pages). I'd intended to pull together the Orphan Works essay a couple of months ago, but life got in the way.

Maybe you'll get a light and fluffy issue some time this fall. Maybe not. Your feedback is always welcome, even though I don't use it as often these days (and haven't received as much).

Investigating the Biblioblogosphere

What's going on in the biblioblogosphere? I hate the term, but it's convenient. Jon Garfunkel at Civilities (civilities.net) gave me the idea with his "Social Media Scorecard" and related posts—but this isn't directly comparable to his evaluation of 25 "online political writers." Instead, this is an informal study of *a* "top 50" library-people blogs, including some metrics.

Identifying the Candidates

Note that qualifier: *a* "top 50", not *the* top 50—and there are actually 60 blogs in the full investigation, which began with more than 230 candidates. Here's the initial process, as noted in *Walt at Random*:

➢ The "list of candidates" came from three obvious sourcesy: LISFeeds, the Open Directory LIS Weblogs and Collaborative Weblogs subdirectories at Open Directory, and the Libdex Library Weblogs list.

➢ I clicked to each blog (printing out lists to avoid duplication), clicked on "Sub with Bloglines" on the FireFox Bloglines toolbar, and subscribed to up to three of the most general feeds (if there were multiple feeds).

➢ I reset Bloglines to show all listings, clicked on each feed, added up the numbers, jotted down the total on the printed list, then unsubbed all but one feed for each blog.

Baseline criteria for inclusion: Blogs by one or a small group (up to four) of self-identified library people (not "official library" blogs or large-group blogs such as PLA Blog and LISNews), with at least one posting in 2005 (some of the lists don't weed dead blogs) and at least one RSS/ATOM feed (because it's too hard to investigate otherwise).

The first pass resulted in 238 blogs, narrowed to 231 during initial testing.

Narrowing the Field

The next step was to consider "reach" beyond apparent Bloglines readership, particularly because Bloglines numbers can be tricky.

The usual metric for blog reach is the Technorati result. I didn't focus on that, partly because I found it difficult to keep Technorati going long enough to do

many searches at once, partly because I think Technorati overvalues blog linking and, specifically, blogrolls. I did get Technorati linked-site numbers for the 60 "finalists" in the study, and only a few of them vary much in rank from the Reach score I prepared.

I did "link:" searches on three major open-Web search engines that support such searches: Google, MSN and AllTheWeb (one of several Overture-owned sites using the same search engine).

The results were startlingly varied, although blogs at or near the top of the list in one measure tended to stay there in others. But the numbers! For *The Shifted Librarian*, Google showed 9,430 links; MSN showed 76,676; and AllTheWeb showed *449,000*! Totaling all 231 sites, here's how extreme the differences were:

➢ Bloglines showed a total of 31,636 feeds (which I extrapolated to 126,544 readership).
➢ Google showed a total of 71,401 links.
➢ MSN showed a total of 297,345 links.
➢ AllTheWeb showed a total of 2,295,436 links.

While links tended to be proportional across search engines, that wasn't always the case. I could suggest at least half a dozen rankings based on that set of raw data. In the end, I calculated a "Reach" figure based half on apparent direct readership (Bloglines times 4), half on a weighted average of reported links, where weighting was proportional to the overall numbers:

Reach = (4 * Bloglines) + (0.67 * Google) + (0.16 * MSN)+(0.02 * AllTheWeb).

The spreadsheet for this analysis is available at http://waltcrawford.name/liblograw.xls. Feel free to manipulate the calculations to arrive at your own rankings and draw your own conclusions.

Results and Metrics

Any metric such as Reach, which runs from 6 to 47,602 for a group of 231 items, will have obvious breaks in the sequence—gaps that can be used to separate groups of entries. I was looking for 50 to 60 blogs for further analysis. One of several gaps was in that range. After eliminating a few blogs (for reasons noted below), there were 48 blogs in the "top 50."

I added blogs that, while not scoring in the top 50 on Reach, were either in the top 40 of Bloglines subscriptions or the top 30 in Google links, MSN links, or AllTheWeb links. Any of these—and, arguably, others as well—could belong in the "top 50."

For the brief writeups that end this Perspective, I've arranged blogs by descending Reach within three broad groups. The first group has the broadest reach, with a Reach in excess of 2,000. The second group includes the remainder of the "top 48" based on Reach. The third group includes those that score high on specific measures but lower on Reach.

A few blogs that made it through the initial filter were removed from the final group, either because they're large-group blogs, I could not reach the blog on several tries, or there have been no posts since March 31, 2005. Five candidates disappeared for those reasons. In two other cases the design of the blog was such that I couldn't do reasonable metrics. I was able to prepare metrics for 60 blogs.

Metrics

A word about median and mean, for those who haven't dealt with statistics much lately. What's usually called the "average" is the *mean*—the sum of all figures for a given measure divided by the population. It can be a wildly misleading figure. For example, the mean or "average" AllTheWeb link count for the 231 blogs was close to 10,000—but only 38 of the 231 blogs had at least that many links. Given that level of heterogeneity, I also calculated the *median*, another form of "average": the value at which half of the population will be higher and half lower.

I noted the following for each blog, based only on internal evidence:

➢ **Starting date (from archives or About page).** One blog began in 1998, two in 1999, and two in 2001 (none in 2000). Nine began in 2002, 27—almost half of the group—in 2003, and 12 in 2004. Seven new blogs (begun January 2005 or later) have already established reasonable reach.
➢ **Frequency of posts, April-June 2005.** I counted them, going through each archive (except for blogs with numbered posts). The "average" is 85 posts; the median is 58.
➢ **Total length of posts, April-June 2005.** I measured using Word's word count; in some cases, sidebars were included. For ten blogs, where the archives collapse all or most of each entry, I assigned an arbitrary length based on 100 words per post, knowing that this is a relatively low number. Mean total words for that quarter: 14,852 words; median: 11,640 words.
➢ **Average length of posts, April-June 2005.** Which bloggers are essayists and which ones

believe in terse comments? For ten blogs, there's no real number. The mean length across all 60 blogs is 250 words per post, just about a screen's worth; median is 188.

➢ **Number of comments, April-June 2005.** Eighteen of the 60 blogs don't support comments or didn't have any during this period. That helps explain the gulf between the mean (48 comments) and the median (22).

➢ **Comments per post, April-June 2005.** Discounting 18 blogs with no comments, this is a measure of the conversational intensity of the blog. There's a substantial gap between the mean (0.85 comments per post) and the median (0.58).

➢ **Blogroll on home page**. Only 32 of the blogs have blogrolls on the home page (I didn't require the word "blogroll," just a set of links to other blogs.) Mean: 23. Median: 10.

➢ **Link-based postings.** Most blogs, within libraryland and elsewhere, involve links in most posts, although the posts may or may not be link-based. I looked at the current home page, or the first 10 posts on that page; these are such crude measures that I hesitate to say much about them.

➢ **Technorati sites as of July 25-26**. For what these are worth, here they are. I couldn't get consistent technorati link numbers at all. Mean: 98. Median: 49.

➢ **BlogPulse rank as of July 25**. This number is so dependent on activity during July and the very end of June that it's somewhere between meaningless and misleading in terms of the reach and readership of any blog. Only 21 of 60 blogs had any BlogPulse rank at all—and those 21 leave out six of the 20 blogs with the broadest reach.

An expanded spreadsheet for these 60 blogs, including all the numbers shown above (and the BlogPulse citation number), along with the URL for each blog, is at http://waltcrawford.name/liblog60.xls.

Category Standouts

These aren't necessarily leaders, but they're significant exceptions from the norm.

Starting Date

The oldest blogs in the study—those started prior to August 2002—are ResearchBuzz, Aug-98; oss4lib, Feb-99; librarian.net, Apr-99; LibraryPlanet.com, Jan-01; Scholarly Electronic Publishing Blog, Jun-01; The Shifted Librarian and Infomusings Blog, Jan-02; Caveat Lector and Catalogablog, Mar-02; eclectic librarian, Jun-02; and The Aardvark Speaks, Jul-02.

Frequency (April-June 2005)

Most frequent postings (at least twice the median):

beSpacific	723
Library Stuff	291
Collecting my Thoughts	233
The Ten Thousand Year Blog	211
LibrarianInBlack	206
Phil Bradley's Blog	175
Tame the Web: Libraries and Technology	169
librarian.net	163
ResearchBuzz	142
Beyond the Job	141
Professional-Lurker: Comments by an academic in cyberspace	135
It's all good	120
Caveat Lector	120
scitech library question	117
LibraryPlanet.com	117
Catalogablog	116

Total length of posts (April-June 2005)

The wordiest bloggers—those with total text longer than the mean length:

Collecting my Thoughts	61,823
Professional-Lurker: Comments by an academic in cyberspace	44,460
Library Stuff	39,858
Free Range Librarian	38,093
beSpacific	37,227
Tales from the "Liberry"	35,664
The Shifted Librarian	34,682
Walt at Random	34,379
Caveat Lector	33,902
Beyond the Job	32,262
It's all good	28,561
The Invisible Library	25,439
LibrarianInBlack	21,425
The Ten Thousand Year Blog	21,100
Tame the Web: Libraries and Technology	20,878
Pop Goes the Library	20,292
librarian.net	19,889
The Distant Librarian	18,464
A Wandering Eyre	17,885
TangognaT	16,983

Average length of posts (April-June 2005)

The essayists—those whose posts average longer than the mean length:

LibraryCog	2,362
Scholarly Electronic Publishing Blog**	578

Tales from the "Liberry"	532
Feel-good Librarian	482
Walt at Random	446
The Days & Nights of the Lipstick Librarian!	400
The Invisible Library	391
Pop Goes the Library	383
Free Range Librarian	381
The Shifted Librarian	361
Professional-Lurker: Comments by an academic in cyberspace	329
The Pod Bay Door	306
Confessions of a Mad Librarian	295
Mamamusings	288
TechnoBiblio	284
Caveat Lector	283
Library clips	278
Collecting my Thoughts	265
LawLibTech	255

I would originally have suggested that the cutoff for "essays" was twice the median, or at least 376 words (the first seven blogs in this list), but I think that leaves out too many blogs where posts are a *mixture* of shorter items and relatively long essays. The "**" next to Scholarly Electronic Publishing Blog is because those aren't essays—they're fortnightly entries consisting of many links and brief descriptions.

The terse posters are:

beSpacific	51
Phil Bradley's Blog	57
blogwithoutalibrary.net	64

Total comments (April-June 2005)

These blogs had the most conversations in all, with a clear gap between the least of these and the next highest (just above the mean):

The Aardvark Speaks	255
Collecting my Thoughts	238
Walt at Random	225
TangognaT	175
Free Range Librarian	172
tinylittlelibrarian.blog-city.com	159
LibrarianInBlack	146
It's all good	142
Mamamusings	131
A Wandering Eyre	109
Feel-good Librarian	94
The Invisible Library	66
LibraryLaw Blog	66
The Pod Bay Door	65
The Distant Librarian	62
Tame the Web: Libraries and Technology	60
Pop Goes the Library	60
Librarian Avengers	60

Conversational intensity (April-June 2005)

Given the relatively low rate of comments in general, I've included every blog that averaged at least one comment per post during this quarter:

LibraryCog	3.60
Feel-good Librarian	3.48
Librarian Avengers	3.16
The Aardvark Speaks	3.11
mamamusings	2.98
Walt at Random	2.92
tinylittlelibrarian.blog-city.com	2.37
The Pod Bay Door	2.32
TangognaT	1.97
Information Wants To Be Free	1.95
walking paper	1.81
blogwithoutalibrary.net	1.75
Free Range Librarian	1.72
LibraryLaw Blog	1.40
It's all good	1.18
Confessions of a Mad Librarian	1.15
Pop Goes the Library	1.13
blogdriverswaltz.com	1.06
Collecting my Thoughts	1.02
Library Web Chic	1.02
The Invisible Library	1.02
A Wandering Eyre	1.01

Blogroll

These blogs had 79 or more sidebar links, with the first six having more than 100 each: tinylittlelibrarian.blog-city.com, Professional-Lurker: Comments by an academic in cyberspace, Open Stacks, Librarian's Rant, commons-blog, Collecting my Thoughts, The Ten Thousand Year Blog, The Invisible Library, scitech library question.

Link-based postings: those who don't usually link

This is an awfully small sample, but these five blogs had outbound links within fewer than half of the first ten entries on their home pages: Caveat Lector, Librarian Avengers, The Aardvark Speaks, Feel-good Librarian, Tales from the "Liberry."

Technorati site count (July 25-26)

These blogs showed more in-linking sites than the mean, but given that one well-read blog (LibraryCog) had no Technorati count at all, I question the significance of this measure:

The Shifted Librarian	1,019
librarian.net	540
Mamamusings	424
Library Stuff	403
ResearchBuzz	343
Collecting my Thoughts	294

beSpacific	285
The Pod Bay Door	155
Free Range Librarian	135
Catalogablog	133
Caveat Lector	118
Tame the Web: Libraries and Technology	109
The Aardvark Speaks	101

Blogpulse rank on July 25

Here's the full list of those showing up in Blogpulse:

The Shifted Librarian	359
beSpacific	745
Library Stuff	1,148
LibrarianInBlack	1,308
librarian.net	1,416
ResearchBuzz	1,418
It's all good	1,702
Caveat Lector	1,999
Mamamusings	2,229
Free Range Librarian	2,503
Library clips	3,343
Catalogablog	3,810
The Distant Librarian	4,261
Open Stacks	4,591
LawLibTech	4,995
The Aardvark Speaks	5,255
Walt at Random	5,398
blogwithoutalibrary.net	5,721
blogdriverswaltz.com	6,729
scitech library question	6,753
Professional-Lurker: Comments by an academic in cyberspace	7,908

Brief Comments on Individual Blogs

The blogs below are listed in descending Reach order as measured in late July 2005. I was tempted to list them in alphabetic order, but that seemed silly. For each blog, I've provided the motto or subtitle (if there is one), the mission (if there is one), the author *as named in the blog*, and my own impression of the overall mix of professional, topical, and personal postings. My note about "voice" has to do with how much you "hear" the blogger in these entries—whether they're neutral in tone or like listening to the author. I add a note about interesting metrics if any are distinctly above average or, for average post length, below average—but you should recognize that what makes a blog worthwhile is not its metrics but its content.

Where there are categories with counts of posts for the categories, I list up to five of the most frequently used categories in descending order. If there's no list of categories it's because there aren't categories or the categories don't include counts.

Blogs with the Broadest Reach: Group 1

I'm not going to call these bloggers the "A list" because I don't believe that term is helpful. This group reaches fairly far down the "Power Law," such as it is, with a 23:1 ratio in "Reach" between the first and the last in the group.

The Shifted Librarian

Motto: Shifting libraries at the speed of byte! Mission: One-page discussion, "What is a shifted librarian?" Author: Jenny Levine. Mix of technology, library, and personal entries, strong voice.

Metrics: *By far* the broadest reach by any measure: #1 in every link search. #1 BlogPulse rank (359), the only four-digit Technorati site count (1,019). Many posts, essay-length posts, seventh-highest overall word count. Relatively old (January 2002).

Library Stuff

The start date is probably wrong: Earlier archives have disappeared.

Motto: The library blog dedicated to resources for keeping current and professional development. No mission page. Author: Steven M. Cohen. Primarily topical, some personal; mild voice.

Metrics: Second most frequent poster (291), third highest word count, third highest BlogPulse ranking, fourth highest Technorati site count (403).

ResearchBuzz

Long-established specialty blog. Motto: Search Engine News and More Since 1998. Mission: "ResearchBuzz is designed to cover the world of Internet research. To that end this site provides almost daily updates on search engines, new data managing software, browser technology, large compendiums of information, Web directories -- whatever. If in doubt, the final question is, "Would a reference librarian find it useful?" If the answer's yes, in it goes!" Author: Tara Calishain. Almost entirely professional resource postings, limited voice.

Metrics: Frequent posts, fifth highest Technorati site count, sixth highest BlogPulse ranking. The oldest blog in the study (August 1998).

librarian.net

No current motto; one-page discussion includes mission. Author: Jessamyn West. Professional and some personal posts with strong voice.

Metrics: Frequent posts, fairly high overall word count, second highest Technorati site count (540),

fifth highest BlogPulse rank. Very old blog (April 1999).

beSpacific

Motto: Accurate, focused law and technology news. Mission: "beSpacific focuses on the expanding resources in the public and private sector related to law and technology news. Daily postings provide updates on issues including copyright, privacy, censorship, the Patriot Act, ID theft, and freedom of information." Author: Sabrina L. Pacifici. Primarily resource listings with comments; subdued voice. (Pacifici also founded and runs *LLRX*.)

Metrics: Most frequent postings (723), *shortest* average post length (51 words), fifth highest overall word count; second highest BlogPulse rank (745), seventh highest Technorati site count.

mamamusings

Current motto: elizabeth lane lawley: yet another proud member of the reality-based community. No mission statement. Author: Elizabeth Lane Lawley. Varied academic, topical (social software), and personal posts; strong voice,

Top categories by number of posts give a sense of Lawley's range: technology, on blogging, teaching, travel, idle thoughts.

Metrics: Fifth most intense conversations (2.98 comments per post), ninth most comments overall; longer than average posts; third highest Technorati site count (424).

Free Range Librarian

Motto: Just-in-time librarianship, from K.G. Schneider. The About page implies a "mission" based on Schneider's work and life. Mix of professional and personal posts with strong voice.

Metrics: Fourth largest overall word count, fairly long posts; fifth most comments overall, fairly high conversational level; more than 100 Technorati sites.

Tame the Web

Subtitle: Technology & libraries (or "libraries and technology" in the metadata title). Mission: "The blog includes topics such as current and future technology uses in libraries, training tips, staff development and various other interests concerning library settings." Author: Michael Stephens, with resume and other info directly from home page. Mostly professional and topical posts, including personal travels on professional business; strong voice.

Most frequent categories: librarians, libraries & the profession; top tech trends; LIS blogs rule!; instant messaging & chat; pursuing the PhD.

Metrics: Frequent posts and comments, high overall word count (but relatively brief posts), more than 100 Technorati sites.

LibrarianInBlack

Motto: resources and discussions for the "tech-librarians-by-default" among us… Mission: "I hope this site can serve as a one-stop-shop for all us Techie Librarians...web design, technology news, library world news, reference stuff, funky gadgets, and other useful (or simply amusing) sites and posts." Author: Sarah Houghton, who chooses not to use it on the blog. Mix of professional and personal posts; moderate voice.

Metrics: Fifth most frequent poster, seventh highest comment count, high overall word count (but relatively brief posts), fourth highest BlogPulse rank.

Catalogablog

Subtitle: "Library cataloging, classification, metadata, subject access and related topics." Author: David Bigwood. Primarily focused professional posts strongly related to the subtitle, with subdued voice.

Metrics: Frequent posts, more than 100 Technorati sites. Relatively old blog (March 2002).

commons-blog

Mission: "commons-blog is an American Library Association-sponsored site collecting news, discussion, and commentary related to the information commons in theory and practice, along with announcements of updates to the info-commons.org main site." While the editor is Frederick Emrich, most recent posts come from "misseli" (see Confessions of a Mad Librarian). Mostly topical posts with a fair amount of personal commentary and opinion, moderate voice.

Caveat Lector

Motto: Reader beware! No mission statement. Author: Dorothea Salo. Mixture of personal, professional, and topical posts, with *very* strong voice. Uses Latin dates and division headings on the page.

Metrics: Frequent posts, ninth highest overall word count, longer than average posts, eighth highest BlogPulse rank. Relatively old (March 2002). Generally self-contained posts (few outlinks).

TechnoBiblio

Current motto: Technology + Libraries = Here. Mission (in part): "TechnoBiblio was originally created to

be a resource where librarians could keep up on technology news. Since its first post in May 2003 the scope has expanded to include news and opinions that relate to technology and any part of the information science realm."

Group blog by four authors, three of them currently or formerly associated with the Gates Foundation. Mostly professional with some personal comments; mild voice.

Metrics: Longer than average posts.

Lorcan Dempsey's blog

Motto: On libraries, services and networks. No stated mission. Author: Lorcan Dempsey. Professional postings with personal commentary; assured voice.

The first single-writer blog from OCLC, and the first with OCLC's logo.

The Aardvark Speaks

Motto: Essence, effervescence, obscurity. Mission: There is a blog manifesto, but you'll have to read it yourself. Author: Horst Prillinger (a librarian, teacher and university lecturer in Vienna), although the "about" page also lists Haldur Gislufsson (a moose) and Richard Ellenson. The only European blog on this list, as far as I know. Mostly personal with some professional posts; vivid voice.

Metrics: Most comments (255) and fourth most extensive conversations (3.11 comments per post). Slightly more than 100 Technorati sites. Relatively old (July 2002). Self-contained posts (few outlinks).

Open Stacks

Motto: Promoting information access and literacy for all. No mission statement. Author: Greg Schwartz. Most current postings relate to Schwartz' podcasts, but there are other professional and personal posts; strong voice.

Most frequent categories: bloggery, podcasting, news, commentary, tangent.

SiteLines

Subtitle: Ideas about web searching. Mission: "SiteLines is intended to present a distillation of the most important trends, news, and new web search tools and directories." Author: Rita Vine. Primarily topical posts related to web searching, with authorial commentary and opinion as appropriate.

Most frequent categories: Google, resources—misc., search engines—business issues, resources—health, searching—best practices.

blogwithoutalibrary.net

Subtitle: a blog about what libraries are doing with blogs, rss, & other little technologies. Mission: "At bwal.net, you will find discussion about and links to some of the very interesting and engaging ways libraries are making use of blogs, RSS, and other emerging technologies to serve their users." Author: Amanda Etches-Johnson (who *is* in a library, McMaster University). Mostly topical postings with appropriate personal commentary.

Metrics: Third shortest posts (64 words), fairly extensive conversations.

walking paper

Mission: "the term 'walking paper' is a way in which i think of new(ish) information technologies. things like IM and text messaging are like active, animated paper to me. the term also is one letter away from 'walking papers,' something that librarians don't want to be handed (and shouldn't be handed, if we pay attention) because of technology." Author: Aaron Schmidt. Mix of professional and personal posts; strong voice.

Metrics: Extensive conversations (1.81 comments per post).

scitech library question

Motto: Occasional postings of interest to engineering and scitech librarians. Also known as STLQ. Group site, primarily Randy Reichardt with contributions from three other science librarians. Generally professional, topical posts with some commentary.

Long list of specific categories, with these most common: publishers & publishing, in the news, open access, dbs & db producers, scholarly publishing.

Metrics: Frequent poster.

Blogs with Fairly Broad Reach: Group 2

LibraryPlanet.com

No motto or mission, possibly because the site seems partly broken. No stated author, probably because the "About" page isn't there. Fairly strong authorial voice; mostly professional posts with commentary.

Metrics: Frequent poster. Old blog (January 2001).

The Days & Nights of the Lipstick Librarian!

Motto: The diary of a library fashionista…. Author: "absherl," or the Lipstick Librarian. Very strong voice. More professional posts than you'd expect given the blog name, with thoughtful commentary.

Metrics: essays (400 words per post).

It's all good

Mission: "A blog from 3 OCLC Online Computer Library Center staff about all things present and future that impact libraries and library users. A conversation that starts with the Environmental Scan and wanders around from there." Group blog, all posts signed. Strong voices, mix of professional and personal posts with commentary. Article count does not include many one-photo postings. The first publicly-visible OCLC blog, but not hosted on or vetted by OCLC.

Metrics: frequent posts, eighth highest overall comments, extensive conversations, high overall word count, seventh highest BlogPulse rank.

The Invisible Library

Motto: The Blog Of An Open Source Librarian, In Which There Is No Shushing. No mission statement. Author: "Keith." Strong voice; personal, political, and sometimes professional posts. Unabashedly liberal and an open-source advocate.

Metrics: many comments, extensive conversations, high overall word count, essay-length posts.

The Ten Thousand Year Blog

Motto and mission: Archivist-historian David Mattison's musings and Web tracks on digital culture preservation issues. Author: David Mattison. Mostly wide-ranging professional/topical posts with light voice, but Mattison's willing to express a strong and thoughtful opinion when it's called for. Word count assigned (archives collapse long posts).

Most frequent categories: *Searcher* magazine threads, information knowledgists, digital libraries and collections, digital preservation, history findings.

Metrics: Fourth most frequent poster.

Library Monk

Subtitle: The blog of Dan Greene. Mission: "Library Monk is: Thoughts of Dan Greene on Library and Information Science, Information Technology, web design, life, religion and perhaps monks." Author: Dan Greene. Personal, wide-ranging topical, sometimes professional posts, with strong voice.

Most common categories: thoughts on life, of interest online, information technology, library monk news, libraries.

Library Web Chic

Motto: Resources for librarians who are interested in the application of web design and technologies in libraries. No mission statement, empty About page.

Author: "Coombs, K. A." based on articles and presentations page. Wide-ranging mix of professional and personal posts with solid (heavy first person) voic.

Most frequent categories: general thoughts, notes from the field, library systems, CSS, usability. (Note that "general thoughts" has twice as many entries as all other categories combined.)

Metrics: extensive conversation.

Confessions of a Mad Librarian

Subtitle: A forum for discussion of library and information topics and interests by a rank amateur, a dilettante and a gadabout. No mission. Although by no means anonymous, author identified only as "misseli" in posts (no About page). Wide-ranging, mix of professional and personal, strong voice.

It's worth pointing out that Eli Edwards, misseli, has been the primary author at commons-blog for some time.

Metrics: Extensive conversations, longer than average post length.

TangognaT

Motto: I'm a bibliomaniacal palindrome! No mission, quixotic About page. Pseudonymous. Mix of personal and professional entries (with strong emphasis on anime) with strong voice.

Metrics: fourth-highest number of comments, extensive conversations.

Walt at Random

Motto: Libraries, music, net media, cruising, policy, and other stuff not quite ready for Cites & Insights. No mission. Author: Walt Crawford. Odd mix of personal, randomly topical, and (rarely) professional postings; strong voice.

Most frequent categories: net media, libraries, writing and blogging, stuff, movies and TV.

Metrics: third-highest number of comments, sixth most extensive conversations (2.92 comments per post), fifth-longest essays (446 words), eighth-highest overall word count. Young blog (April 2005).

oss4lib

Motto: open source systems for libraries. "Our **mission** is to cultivate the collaborative power of open source software engineering to build better and free systems for use in libraries. Toward this end, we maintain a listing of free software and systems designed for libraries (the physical, books-on-shelves kind), and we track news about project updates or related issues of interest." Author: Dan Chudnov and

various volunteers. Nearly dormant topical blog with light voice.

eclectic librarian

Motto: Links, commentary, and other musings by a serialist. No mission or About page. Author: Anna Creech (each post signed in full). Mix of personal and professional postings, with moderate voice. Word count assigned: archives collapse posts.

Most frequent categories: library, blog, political, it's all about me, books.

Metrics: relatively old (June 2002).

LibraryLaw blog

Subtitle and mission: Issues concerning libraries and the law - with latitude to discuss any other interesting issues Note: Not legal advice - just a dangerous mix of thoughts and information. Brought to you by Mary Minow, J.D., A.M.L.S. [California, U.S.] Author: Mary Minow (and three contributors). Mostly professional, with voice as appropriate. Word count low: Some very long posts partially hidden. One of the premier posts relating law and librarianship.

Metrics: many comments, active conversations.

Collecting my Thoughts

Subtitle: Essays, stories, poetry, memories, comments on the news. No mission or About page. Author: "Norma." (Norma Bruce, not particularly hidden.) Almost entirely personal and right-wing political, with very strong voice.

Metrics: Highest overall word count (61,823 words), third most frequent poster, second most frequent comments, active conversations.

Phil Bradley's Blog

Subtitle and mission: "For librarians and people interested in search engines, searching the net, designing web pages, new utilities and so on. Short commentaries, keeping you up to date with the world of internet search and design." Author: Phil Bradley. Almost entirely focused news items with light voice; largely brief items (the word count is correct).

Metrics: Frequent poster, second shortest posts (57 words).

BlogJunction

Motto: WebJunction's niche in the blogosphere. Group blog, four authors. Mix of topics, largely WebJunction-related, with fairly strong voices. Word count assigned: Collapsed archives.

Most frequent categories (of six total): online collaboration, news, library services, tech.

Metrics: young blog (April 2005).

Librarian Avengers

No motto (except on a linked page, "our metadata can beat up your metadata." Mission: "Librarian Avengers started in 1998 with an essay titled 'Why you should fall to your knees and worship a librarian.' I was working in a digital library at the time, and the overwhelming response from the amazing and warm library community was one of the things that propelled me into grad school. I blogged grad school (I promise to put the archives up someday), made some t-shirts, and now there's this." Author: Erica Olsen. Mix of professional and personal entries, with strong voice.

Metrics: Third most extensive conversations (3.16 comments per posts), frequent comments. Generally self-contained posts (few outlinks).

Beyond the Job

Mission and author/editors: "Professional tips for librarians: Articles, job-hunting advice, professional development opportunities, and other news and ideas on how to further your library career. Compiled by the Library Job People, Sarah Johnson and Rachel Singer Gordon." Almost entirely notices of events and opportunities, with very little voice or commentary.

Metrics: Frequent posts, tenth-highest overall word count.

ONLINE Insider

Mission and author: "*Welcome to Online Insider … the editorial blog by Marydee Ojala, Editor of ONLINE: The Leading Magazine for Information Professionals.* ONLINE Insider intends to extend the reach of the print publication, presenting a more timely commentary on the products, people, and events that shape today's online world. It explores new technologies as they impact the working lives of information professionals, explains resources for specific topic areas, and expounds on information management tools and techniques." Primarily topical with moderate voice. Surrounded by Information Today stuff.

Metrics: young (January 2005).

The Information Literacy Land of Confusion

Mission and author: "Blog of librarian Michael Lorenzen discussing library user education, library instruction, librarianship, information literacy, and search engines. Also covers other observations on life in general." Much as stated; fairly strong voice.

A Wandering Eyre

Subtitle: A bibliophile's musings on books, libraries, the world, life, and anything else that comes to mind. Author: "Jane." Mix of personal and professional, with strong voice.

Metrics: Tenth highest number of comments.

Scholarly Electronic Publishing Blog

Mission contained within the *Scholarly Electronic Publishing Bibliography*, for which this serves as an update/announcement service. Fortnightly postings with extensive lists of citations relevant to the topic. Author: Charles W. Bailey, Jr.; light voice.

Metrics: Second longest posts, but that's misleading: Each "post" has many individual subposts. Old blog (June 2001).

LibraryCog

Motto: it's like, you know, library systems and stuff.... Author: Art Rhyno, although that's not evident on the blog. Professional bent with strong voice, very few, very long essays.

Metrics: Most extensive conversations (3.6 comments per post), longest essays (2,362 words per post). Young blog (January 2005).

Feel-good Librarian

Motto: Why we do this. Author not stated. Mostly essays about life in the library; strong voice.

Metrics: Second most extensive conversations (3.48 comments per post), fourth longest essays (482 words per post). Young (February 2005). Generally self-contained posts (few outlinks).

the pod bay door

No subtitle or stated mission. Author: Randy Reichardt. Wide-ranging posts with strong voice.

Most frequent categories: film, miscellaneous, random thoughts, music, blogging.

Metrics: frequent comments, eighth most extensive conversations (2.32 comments per post), relatively long posts, more than 100 Technorati sites.

Information Wants To Be Free

Subtitle: A librarian and tech geek reflecting on the profession and the tools we use to serve our patrons. Author: Meredith Farkas. Mix of professional and personal posts with strong voice. The starting date and other counts may be off: "Previous posts" resulted in a 404, and there are no archives.

Most frequent categories: our digital future, libraries, job search, random, blogging.

Metrics: extensive conversations.

Pop Goes the Library

Subtitle: An ongoing exploration of the intersection of popular culture and libraries. Mission: "We're public librarians. We believe libraries can learn from and use Pop Culture to improve their collections, services, and public image. We love TV, music, the movies, comic books, anime, magazines, all things Net...you get the picture." Group blog (three authors). Most posts relate media to libraries, but with some additional range. Strong voices.

Metrics: frequent comments, fairly extensive conversations, higher-than-average overall word count, fairly long posts.

blogdriverswaltz

Mission: "I like the idea of being able to share my findings, ideas and opinions about librarianship, life and everything in between in an open forum that encourages dialogue. This is blogging - plain and simple. It's fun to have a voice in a non-traditional form of publishing. A timely form of publishing. A form of publishing subjected to a different sort of peer review than we might traditionally think of in academic conversation." Author: Geoff Harder. Mix of posts, strong voice. Start date unclear.

Metrics: extensive conversations.

Broad Reach by Some Measures: Group 3

I've included here blogs that, while not scoring in the top 50 (well, 48) on the final "reach" measure, were in the top 40 of either Bloglines subscriptions or the top 30 in Google links, MSN links, or AllTheWeb links. Any of these—and, arguably, others as well—could belong in the "top 50."

Librarian's Rant

Motto: Planning the Revolution… Mission and author: "*The Librarian's Rant* is the steam let off by Louise, a public reference librarian lost in the Heartland." Wide mix of posts, strong voice.

LibraryTechtonics

Motto: A librarian riding the shift. Mission: "The long-term plan for this site will include a resume, portfolio, and hopefully all kinds of other fun, including something I'm brewing in the back of my head as 'Extreme Reference Challenge.' More on that later, so keep coming back." Author: Andrea Mercado. Mix of professional and personal posts, moderate voice.

Tiny Little Librarian

Subtitle: Musings of a too-short girl in the high-stacks game of librarianship... Very anonymous. Mostly posts about life in the writer's library, with some other professional and personal posts; strong voice. Word count is assigned: archives collapse text.

Metrics: sixth most frequent comments, seventh most active conversations (2.37 comments per post).

The Distant Librarian

Subtitle: Comments on the world of Distance Librarianship. Author: Paul R. Pival. Focused blog, primarily posts on distance librarianship, moderate voice.

Metrics: frequent comments, essayist.

Professional-Lurker

Subtitle: comments by an academic in cyberspace. Author: Louis Ann Scheidt. Mostly professional, moderate voice.

Metrics: frequent posts, second-highest overall word count, essayist.

dave's blog

Subtitle: Cool stuff about library web sites - usability, searching, new technologies, design ... and whatever else I decide to post :-). Author: David King. Mostly professional posts with moderate voice.

The Laughing Librarian

No subtitle. Mission statement: "The Laughing Librarian website just *is*, okay?" (With an excellent gloss.) Author: Brian Smith, "who is **not** 'The Laughing Librarian'." Very much on topic with moderate voice.

Tales from the "Liberry"

Subtitle and mission: "An employee of a small town 'liberry' chronicles his eternal quest to remain sane while dealing with patrons who could double as extras in a David Lynch film." Author: "Juice," anonymous for good reason. Most posts are vivid essays about the users of a small West Virginia library, with a few personal posts from the author's own life.

Metrics: Sixth-highest overall word count, third-longest essays (532 words per post). Generally self-contained blog (few outlinks).

infomusings blog

No subtitle or mission. Author: "katrina." When this blog is active—it varies widely—it's a wide mix of library, "infomusings," and personal entries, with moderate voice.

Metrics: relatively old (January 2002).

Library clips

No subtitle. Mission: "This blog is a space for me to share some ideas, thoughts, and feedback…from a librarian's point of view." Author: "johnt" Mostly topical posts related to various internet-and-library topics, moderate voice.

Metrics: relatively long posts. Young (February 2005).

Filipino librarian

Subtitle: For those interested in knowing more about the Philippines, Filipiniana, Philippine libraries and Filipino librarians. Motto: "I can't change Filipino librarians,but I can change the way you look at us." Author: "vonjobi." Wide range of library-related posts with a Filipino slant, moderate voice.

Metrics: young (February 2005).

LawLib Tech

Subtitle: A conversation on law library technology and knowledge management. No mission statement. Author: Cindy L. Chick. Primarily topical with moderate voice.

Metrics: relatively long posts.

My Comments

What does it mean? A few points are obvious:

- ➢ Most library-oriented blogs aren't in traditional "A list" categories, although *The Shifted Librarian* comes close.
- ➢ Quite a few library-oriented blogs have fairly broad reach—sometimes outside the library community (again, *The Shifted Librarian* stands out) but mostly within the community.
- ➢ There's no single model for a librarian blog. beSpacific churns out more than two dozen very brief pointers a day; *Tales from the "Liberry"* offers a 500-word commentary on the charms of small library life four or five times a week. *librarian.net* maintains extended conversations of sorts despite not accepting comments directly; several blogs (including mine) are substantially enriched by frequent, thoughtful comments.

How many of these 60 blogs fall into the idle chatter and semiliterate categories attacked by people who should know better? A few are rich in the personal lives of the bloggers, but most aren't. I wouldn't accuse any of these bloggers of lacking writing skills. I would be hard-pressed to choose even a dozen I don't consider worthwhile.

I was reading posts from 231 library-related blogs during part of this evaluation—and although I've been unsubscribing little by little, more than 150 are still there. Most, to be sure, don't post very often. Of the 60 profiled here, there may be six or eight that I'll eventually unsubscribe, but in at least half the cases that's because they focus on an area I don't much care about at the moment.

The Shifted Librarian isn't worth reading *because* it tops all the reach metrics. It's the other way around: it tops all the reach metrics because Jenny Levine cares deeply about "shifting," a whole range of digital technologies, and how they might interact with libraries—and has no qualms about letting you know what she thinks. She's become widely read outside the library community. I may and do disagree with her at times—and when I do, I say so, because she's worth paying attention to.

That's the extreme example. Don't try to take away my feed from *Caveat Lector* or *The Aardvark Speaks* or *LibrarianInBlack* or *It's all good* (or *hangingtogether*, way too new for the study)—those and many other voices enrich my life and my understanding of a range of issues and ideas.

I didn't track gender of bloggers while doing the study, because I didn't much care. But since it seems to be a big deal within some segments of the community, so here's the summary:

Of the 20 blogs in Group 1, 10 are clearly by women, 7 by men, 3 groups. Group 2 is evenly split: 11 each, 6 unknown or groups. Group 3 has five women, four men, three unknown or groups.

Overall, that's 26 women, 22 men, and 12 unknown or groups out of the 60 considered here. You can look at the overall percentage of women in librarianship and say that's low; you can look at the percentage of women in the technological side of librarianship and conclude that it's about right.

I'd love to extend this study to the top *half* of the list, or at least the top hundred, but time doesn't permit that. Sorry, *Tinfoil + Raccoon*, *DigitalKoans*, *Library Dust*, *Travelin' Librarian*, and others—I had to limit the scope of this particular study. Maybe someone else will pick it up. Maybe I'll do another one in a year or two to see how the biblioblogosphere has changed.

If you're unaware of blogs in this particular "top 50," you might try a few of them—both spreadsheets include URLs accurate as of late July 2005. But don't stop there: Sample a few other blogs that sound inter-esting. Most aggregators make it as easy to unsubscribe as to subscribe.

If your favorite blog (or your own blog) isn't here, so what? If it's doing something worthwhile to a few people (or a few hundred people), who cares if it isn't reaching the multitudes?

Perspectives
Summertime Blahs

It's summer, not always a lazy time for me—and this summer's been extraordinarily busy at work and began with other issues. The net effect has been more than the usual summer lassitude. Not so much lassitude that I didn't carry out the "biblioblogosphere" investigation, but enough to discourage Big Essays on Big Topics. So here are a couple of medium-size essays on topics that may (or may not) deserve better.

Iconoclasm and the Great God Google

This one's been simmering for a while and bubbled up during the first half of July thanks to a spirited and sometimes bemusing set of threads on Web4Lib. I used the discussions as a springboard for my "disContent" column in the November 2005 *EContent Magazine*. If you want to see all the text, web4lib's new home is web4lib@webjunction.org. (You may have to join in order to view the archives.) The threads include "Another Google question" and "Google limit of 1,000 results." At least one of them started before July 6, where I picked them up.

Most of you are probably aware of the background. Google is now the most popular open-web search engine, although Google searches are still a minority of web search-engine searches according to most studies. Google has become *synonymous* with web searching, just as AltaVista may have been a few years back—but on a much larger scale as more people regard the web as a fact of life. Because Google's ranking algorithms worked remarkably well for several years, it's gained a stature that none of the other engines currently enjoys—even though, at least in my opinion, the algorithms don't work *nearly* as well as they used to. Google has also acquired or created a number of other services, some still beta, some near-failures (e.g., Google Answers), some successes. This is all great and good, and affects libraries by providing a useful, easy-to-use tool for searching a broad new range of resources.

You can throw in Google Scholar and Google Print if you like—not for what they are now but for how they're perceived. That perception, as for Google itself, is the heart of the problem that resulted in some of the attitudes shown in this discussion.

The perception is threefold, overstated here for clarity:

➤ Google can do no wrong and all Google plans succeed.

➤ Google does right what library systems and services (online catalogs, fee-based indexes) do wrong—and the solution is for library systems and services to be "just like Google." *Not* paraphrasing: "Google is all you need."

➤ For the pessimists, Google dooms libraries, because "everyone" "always" wants to use the simplest means possible to get something "good enough" (and, of course, because Kids These Days don't *really* read print books or use anything that's not digital).

There are related issues, but I'll leave those for later.

One of those raising questions about Google's unlimited and universal excellence for all online tasks is Roy Tennant, about as fierce a critic of traditional online catalogs as you'll find. Roy knows there's loads of room for improvement in library tools, but he's also aware (as anyone working for the California Digital Library should be) that there are many different users with many different needs, and one set of tools won't handle them all equally well. More specifically, Google *doesn't* replace the scholarly apparatus and research databases—not even online catalogs.

I don't recall whether Bernie Sloan or Patricia F. Anderson raised the point first (I think it was Bernie), but they noted what some of us have known for some time: You don't *know* how large a Google result is (except for small ones), because you'll never see more than 1,000 records. For all I know, "about 28,400" records for the phrase "Walt Crawford" (searched on August 8, 2005) could be something over 28,000—or it could be the 975 records I get when I "repeat the search with the omitted results included" (which gives "about 28,600" as an unviewable set). Initially, I can only view 460 records.

Did the denizens of Web4Lib jump all over Google, shouting Gormanesque attacks and asserting that Google should never be used in libraries or for research? Not at all. The harshest attack I could find anywhere in the thread was Roy Tennant's:

Google does one thing, and it appears to do that one thing well. But let's not make the unfortunate assumption that it does more than that one, very specific thing.

Roy's talking about "home Google" here. He's pointing out that Google involves a specific set of assumptions about user needs with no way to change those assumptions. You can't *really* get the most recent pages (and Google's algorithm tends to bury new linking pages), since the date limit is mostly useless (in all web search engines: it's the nature of web page dates)—and you can't examine the entirety of most result sets.

This mild criticism and a slightly more pointed one that follows (related to Google's version of link: results, which has become useless or misleading) are a form of iconoclasm: Suggesting that an icon isn't all that it should be. Nobody—*nobody*—on Web4Lib was saying Google's useless in general.

Battling Iconoclasm

Lars Aronsson said flatly, "no real searchers would be interested in more than the first 900 hits." When offered real-world examples where *researchers* would want to examine an entire result set, Aronsson denied the validity of the examples. He wasn't the only one. Mike Taylor opined:

We *couple of hundred information professionals* on this list care deeply about this stuff, but we do need to come to terms the fact that no-one else does. As far as the other 5,999,999,980 people out there are concerned, Google is just fine. If we pretend otherwise, we're hiding ou[r] heads in the sand.

Aside from the fact that Web4Lib's membership is in the thousands, and that there are at least a couple of hundred thousand librarians in the world who *should* care, Taylor negates several million researchers of all stripes—scientists, lawyers, doctors—for whom "Google is just fine" is an unethical and dangerous stance. (Karen Coyle responded: "The fact that they think it's fine doesn't make it fine. Ignorance may be bliss, but it's a lousy basis for what purports to be an 'information society.'")

Jeremy Dunck was "surprised at the…glee that folks on this list seem to take at poking holes in information tech that's available." I saw no glee. I did see an attempt to discuss real-world flaws and to emphasize that Google is *not* the universal answer to all searching needs. Dunck also questioned the "Google fetish" on a list about "using the web to the benefit of

libraries," an odd challenge given that *everyone* who criticized Google on this list recognizes that it is a powerful web tool to extend reference services.

Here's a sample of that "gleeful" approach from Patricia F. Anderson:

> I like Google just fine, but it is far from doing everything I'd like it to do. I also see no reason for Google to *try* to do everything—some specialized tasks are best in a niche market, where the people who truly care about that will pay attention and take care of it. I don't need or want one search engine that tries to be all things to all people.

That's about as "bashing" an attitude as I saw, unless you count Roy Tennant's continuing insistence that Google doesn't do everything equally well and that researchers sometimes have legitimate reason to want to see items "less relevant" by Google's algorithms.

Why do I call this iconoclasm? Because the critics don't accept Google as an object of veneration; they refuse to treat it as a religion. When you assert that Google *does* answer everyone's real world needs and object to any criticisms, you're treating it as an icon. Iconoclasm is the sensible result.

Placing too much faith in Google? Consider the mixed messages in Ryan Eby's contribution. In the first paragraph, Eby says he uses Google "nearly exclusively (for web search) ever since [around its inception]. In that time I don't think I've ever had a completely failed search…" Eby "never go[es] past two or three pages" and regards wanting deeper results as something "spammers would love." Later, after saying he uses Google nearly exclusively and it always works, Eby comes up with this:

> I personally, and everyone I know, know that Google is not the one stop shop for all research (nor would I want it to be), though it does a damn good job at some things.

Mike Taylor came back with the claim that Google's "big, big win" is that "its top hit (or second, or third) is nearly always the one you want." Taylor has a lot better luck with Google than I do—I'm finding that the results I want are frequently down in the second page these days, with semi-relevant commercial stuff taking up the first page. Karen Coyle's response to Taylor was that Google is "very good at…the retrieval of pages based on proper names…where there is a single obvious answer…. For other types of searches, Google doesn't work so well. There's no 'conceptual' searching." (Names of people also tend to be more difficult as Google's index grows.) Taylor agreed with this observation—but concluded that most people do most of their searching using specific known items. (How he knows this universal truth is beyond me.)

Oddly enough, I'm finding that to be less true as well, particularly in the hospitality area. When I want to find a hotel's website, if I don't know what chain it's part of, Google can be frustrating. Sometimes—maybe half the time, maybe more—the hotel's or resort's website comes out on top. Other times, reservation systems and chambers of commerce and other entities have succeeded in linking their way to the top.

Later, after discussion of a specific problem noted below, Bill Drew raised the "Google bashing" cry, calling it "nitpicking about obscure features." Jennifer Heise suggested that the tone of the discussion "has really come across as 'why Google is bad," as opposed to the continued "why Google isn't the universal solution" that I saw.

Roy Tennant responded:

> I'm trying to more fully understand what Google is good at and what it isn't good at. Given that Google is not very forthcoming on the help pages about limitations such as have been surfaced here by Bernie Sloan and others, this discussion seems to be one of the few places to get such information.

Bernie Sloan also responded: "My motivation is curiosity…trying to find out why things don't seem to work quite as they seem to be advertised…" He also notes cases where he would legitimately go past 1,000 results.

The Link: Problem

The "no more than 1,000, based on our secret relevance sauce" issue is just that: An issue, not a failing. Most other open-web search engines have similar limits. Paid database vendors don't have the luxury of telling users that they'll get "some of" a result—but we're paid, and *should* be held to a higher standard.

A slightly more serious problem arises with Google's link-to searches: "link:[url]," a specific form of search offered by several open-web search engines. Bernie Sloan tried some of these searches and got results that made no sense to him. He asked Google about it and got this answer:

> Our link search does not return a comprehensive set of results. The results will show a sample of the links that point to your page, but this list is in no way indicative of the link structure utilized by Google to formulate a page's PageRank.

Sloan wasn't asking how the PageRank was calculated. He wanted to see who was pointing to one of his own pages. That's reasonable. I've done the same thing. Link counts also enter into any analysis of, for example, blogs (as in this issue's major essay).

Roy wasn't thrilled with this: "So, in other words. The Google 'link:' search is worse than useless. Useless because it fails to work as advertised and worse than useless because it will return just enough so one could image it was working as they supposed (and as depicted by Google)." Andrew Mutch verified that Google's link results "seemed to shrink even as lists of sites known to be linking to the resource had grown." Bernie Sloan suggested that Google should rename "link:" in the explanation: Instead of saying "Find pages that link to the page" it could say, "Find examples of pages that link to the page."

In this case, I don't believe there's a reasonable defense for Google. The link: feature is broken. Google should either turn it off or explain that it's just a sampling.

Nobody's Bashing Google

OK, that's not true. Let's say that nobody on Web4Lib was bashing Google; I can't speak for certain elected ALA officials. What was going on was *librarianship*: Investigating resources to determine when they should and *should not* be used.

The problem here is that too many people see Google as all you need and all you ever will need. That's dangerous. Librarians need to help their users with a broad range of indexes and search tools—and that means understanding limitations of the leaders.

Creative Commons: Foe of Copyright?

Cites & Insights carries a Creative Commons license—one that reserves the right to *profit* from reuse of this material. Want to post this essay on your web? Feel free as long as you're not charging, you cite the original properly, and you note that the essay is protected by copyright. Want to distribute copies to your class? No problem. Want to sell it to others? Big problem unless you ask.

There's now a Creative Commons search engine and Yahoo! searches can include a Creative Commons qualifier. With more than five million CC-licensed websites, that's a good idea. Lots of writers and musicians have concluded that CC licenses make sense, encouraging new creativity while protecting the rights they want protected.

But *any* weakening of maximal copyright, even weakening chosen by a copyright holder, seems to offend some groups. Sometimes, it's a matter of indirection: "My concern is that many who support Creative Commons also support a point of view that would take away people's choices about what to do with their own property," says David Israelite of the National Music Publishers' Assn. (in a May 20 Reuters article, originally from *Billboard*). That's guilt by association, even though CC is precisely a way for people to make "choices about what to do with their own property." Michael Sukin of the International Assn. of Entertainment Lawyers makes a similar leap: "Lessig and his followers advocate a shorter copyright term." True enough—but *entirely* unrelated to CC (which does not lobby for changes in copyright law).

RIAA is not among the CC-bashers: Its president says that artists might *want* to make their music freely available, and that the CC approach solves that need. But there's always a counter-example: Andy Fraser, who wrote "All right now" for Free. He's afraid that he might have used a CC license when he was a young songwriter if one was available, and wouldn't now have the royalties that pay for his AIDS treatment. His solution?

"**No one should *let* artists give up their rights.**" [Emphasis added.]

There it is in a nutshell: You should *not be allowed* to give something away, or even to give it partially away. So much for charity, the public domain, anything other than 100% "I've got mine" capitalism: No one should *let* someone else choose to reduce their own total control over something. This is, in its own way, as totalitarian a statement as any Communist could make, just at the other extreme.

Israelite doesn't go that far, but loves to make broad claims: "Often when people give away their own property under a Creative Commons license, it is really an argument why others should be *forced* to give away their property." I could say that's meretricious bullshit, but "often" is enough of a qualifier to escape such claims. Nobody I know who uses a CC license is making any such argument—and, to be sure, most of us don't "give away" our "property," but grant *some* rights to others while retaining others for ourselves.

Which, according to Andy Fraser, we should not be **allowed** to do. So much for freedom.

Lawrence Lessig wasn't thrilled by the article and placed it in context: *Billboard* has run other pieces es-

pousing this anti-CC view. In an earlier piece, Sukin claimed that CC placed "U.S. copyright income" at risk and included a statement by the writer that CC's "Founder's Copyright" (which they're no longer offering, but which established either a 14 or 28 year copyright in exchange for a $1 payment) was "urging creators to give up their copyright protection" for a buck. It was nothing of the sort, of course; it was a way of establishing a legal contract to limit that protection to 14 or 28 years.

With regard to this particular article, Lessig has a comment on Israelite's assertion as to what CC users are *really* doing:

> I love it when people tell me what my argument 'really' is. The whole premise of Creative Commons is that artists choose. We give licenses to creators. How exactly empowering creators is "really an argument why others should be forced to give away their property" is bizarre to me. By this reasoning, when Bill Gates gives $20,000,000,000 to help poor people around the world, that's an argument for socialism.

Lessig goes on to note that the guilt-by-association link of CC supporters and shorter-copyright supporters is particularly specious: "The RIAA believes it is appropriate to sue kids for downloading music. They're supporters of Creative Commons. Does it follow that Creative Commons supports suing kids for downloading music?"

©2 Perspective
Orphan Works

Orphan works: "copyrighted works whose owners are difficult or even impossible to locate." That's how the Copyright Office defines the term in a January 26, 2005 *Notice of inquiry* on orphan works. That song from 1924 whose writer disappeared or died and whose estate is not clearly identifiable. A photograph taken in 1930. Early documentary films. Tens of thousands of long-out-of-print novels; thousands of pulp-fiction short stories, novellas, and novelettes. The list goes on, literally into the millions—and the inquiry yielded more than 700 direct comments and more than a hundred responses to those comments.

I'm quoting at length from the *Notice of inquiry* because it does a remarkably good job of setting the scene (and, since it's government work, it's in the public domain within the U.S.). Here's the summary:

The Copyright Office seeks to examine the issues raised by "orphan works," i.e., copyrighted works whose owners are difficult or even impossible to locate. Concerns have been raised that the uncertainty surrounding ownership of such works might needlessly discourage subsequent creators and users from incorporating such works in new creative efforts or making such works available to the public. This notice requests written comments from all interested parties. Specifically, the Office is seeking comments on whether there are compelling concerns raised by orphan works that merit a legislative, regulatory or other solution, and what type of solution could effectively address these concerns without conflicting with the legitimate interests of authors and right holders.

Comments were accepted through March 25, with responses through May 9. The "background" section and portions of six "Specific Questions" are excerpted from the *Notice of inquiry*.

Background

The Copyright Act of 1976 made it substantially easier for an author to obtain and maintain copyright in his or her creative works. Today, copyright subsists the moment an original work of authorship is fixed in a tangible form--it need not be registered with the Copyright Office or published with notice to obtain protection. While registration of claims to copyright with the Copyright Office is encouraged and provides important benefits to copyright holders, it is not required as a condition to copyright protection. Under the 1909 Act, renewal registration was required to maintain protection beyond an initial 28-year term. Failure to register the renewal during the last year of the first term resulted in complete loss of protection. The 1976 Act removed the renewal requirement going forward, but kept it for works copyrighted before 1978. It was not until 1992 that the renewal requirement was abolished altogether. These changes, as well as other changes in the 1976 Act and in the Berne Convention Implementation Act of 1988, were important steps toward harmonizing U.S. copyright law with international treaties. Specifically, the Berne Convention and other treaties dealing with copyright that have followed forbid the imposition of formalities as a condition to copyright, principally on the grounds that failure to comply with formalities can serve as a trap for the unwary, resulting in the inadvertent loss of copyright.

Concerns have been raised, however, as to whether current copyright law imposes inappropriate burdens on users, including subsequent creators, of works for which the copyright owner cannot be located (hereinafter referred to as "orphan" works). The issue is

whether orphan works are being needlessly removed from public access and their dissemination inhibited. If no one claims the copyright in a work, it appears likely that the public benefit of having access to the work would outweigh whatever copyright interest there might be. Such concerns were raised in connection with the adoption of the life plus 50 copyright term with the 1976 Act and the 20-year term extension enacted with the Sonny Bono Copyright Term Extension Act of 1998.

The Copyright Office has long shared these concerns about orphan works and has considered the issue to be worthy of further study. On January 5, Senators Orrin Hatch and Patrick Leahy of the Senate Judiciary Committee asked the Register of Copyrights to study this issue and to report to the Senate Judiciary Committee by the end of the year. Also in January, Reps. Lamar Smith and Howard Berman, the chairman and ranking member of the House Judiciary Committee's Subcommittee on Courts, the Internet and Intellectual Property, sent letters to the Register supporting this effort. The Office is gratified that Congress has shown an interest in this important issue and is pleased to assist Congress in its efforts to learn more about the problem and to consider appropriate solutions.

Prior to the 1976 Act, the term of protection was limited to 28 years if the copyright was not renewed. Under this system, if the copyright owner was no longer interested in exploiting the work, or a corporate owner no longer existed, or, in the case of individual copyright owners, there were no interested heirs to claim the copyright, then the work entered the public domain. Of course, it also meant that some copyrights were unintentionally allowed to enter the public domain, for instance, where the claimant was unaware that renewal had to occur within the one year window at the end of the first term or that the copyright was up for renewal. The legislative history to the 1976 Act reflects Congress' recognition of the concern raised by some that eliminating renewal requirements would take a large number of works out of the public domain and that for a number of those older works it might be difficult or impossible to identify the copyright owner in order to obtain permissions. Congress nevertheless determined that the renewal system should be discarded, in part, because of the "inadvertent and unjust loss of copyright" it in some cases caused. More recently, in the mid-1990s, Congress heard concerns that the Copyright Term Extension Act would exacerbate problems in film preservation by maintaining copyright protection for older motion pictures for which the copyright owner is difficult to identify. Also, in our study on Digital

Distance Education published in 1999, the Copyright Office identified several "problems with licensing" that educators asserted in attempting to use copyrighted materials in digital formats, including that "it can be time-consuming, difficult or even impossible to locate the copyright owner or owners."

A situation often described is one where a creator seeks to incorporate an older work into a new work (e.g., old photos, footage or recordings) and is willing to seek permission, but is not able to identify or locate the copyright owner(s) in order to seek permission. While in such circumstances the user might be reasonably confident that the risk of an infringement claim against this use is unlikely, under the current system the copyright in the work is still valid and enforceable, and the risk cannot be completely eliminated. Moreover, even where the user only copies portions of the work in a manner that would not likely be deemed infringing under the doctrine of fair use, it is asserted by some that the fair use defense is often too unpredictable as a general matter to remove the uncertainty in the user's mind.

Some have claimed that many potential users of orphan works, namely individuals and small entities, may not have access to legal advice on these issues and cannot fully assess risk themselves. Moreover, even if they are able to determine with some certainty that there is little or no risk of losing a lawsuit, they may not be able to afford any risk of having to bear the cost of defending themselves in litigation.

Given the high costs of litigation and the inability of most creators, scholars and small publishers to bear those costs, the result is that orphan works often are not used—even where there is no one who would object to the use.

This uncertainty created by copyright in orphan works has the potential to harm an important public policy behind copyright: To promote the dissemination of works by creating incentives for their creation and dissemination to the public. First, the economic incentive to create may be undermined by the imposition of additional costs on subsequent creators wishing to use material from existing works. Subsequent creators may be dissuaded from creating new works incorporating existing works for which the owner cannot be found because they cannot afford the risk of potential liability or even of litigation. Second, the public interest may be harmed when works cannot be made available to the public due to uncertainty over its copyright ownership and status, even when there is no longer any living person or legal entity claiming ownership of the copyright or the owner no longer has any objection to such use.

Empirical analysis of data on trends in copyright registrations and renewals over the last century suggests that a large number of works may fall into the category of orphan works. Based on data of registrations of claims to copyright and their subsequent renewal under the 1909 Act, it appears that, overall, well less than half of all registered copyrighted works were renewed under the old copyright system. Because renewal was required to maintain protection of a work, this data suggests that, at least in many cases, there was insufficient interest a mere 28 years later to maintain copyright protection. The empirical data does not indicate why any particular works were not renewed, and no doubt, a certain portion of those works were not renewed due to inadvertence, mistake or ignorance on the part of the owner. With respect to many of these works, however, particularly those owned by legal entities or other sophisticated copyright owners, it can be assumed that the work no longer had sufficient economic value to the copyright claimant to merit renewal. Libraries and scholars have argued that those works that have so little economic value that they fail to merit the small expense and effort of renewal may nevertheless have scholarly or educational value and should not be needlessly barred from such use.

Specific Questions

1. Nature of the Problems Faced by Subsequent Creators and Users

What are the difficulties faced by creators or other users in obtaining rights or clearances in pre-existing works? What types of creators or users are encountering these difficulties and for what types of proposed uses? How often is identifying and locating the copyright owner a problem? What steps are usually taken to locate copyright owners?

2. Nature of "Orphan works": Identification and Designation

How should an "orphan work" be defined? Should "orphan works" be identified on a case-by-case basis, looking at the circumstances surrounding each work that someone wishes to use and the attempts made to locate the copyright owner? Should a more formal system be established?

…The establishment of a filing system whereby the potential user is required to file an intent to use an unlocatable work has also been suggested. Would the Copyright Office or another organization administer and publish such filings?

3. Nature of "Orphan Works": Age

Should a certain amount of time have elapsed since first publication or creation in order for a work to be eligible for "orphaned" status? If so, how much time?

4. Nature of "Orphan Works": Publication Status

Should the status of "orphan works" only apply to published works, or are there reasons for applying it to unpublished works as well?

5. Effect of a Work Being Designated "Orphaned"

However a work is identified and designated as "orphaned," what would be the effects of such designation? Under systems for a mandatory, formal registry of maintained works, like the 1909 Act, the right to assert one's exclusive rights vis à vis others could similarly be lost, in whole or in part, if the work was not contained on the registry. Should this loss of rights apply only to the particular work at the time of use, or only to the particular use or user, or would it affect a permanent loss of rights as against all uses and users?

6. International Implications

How would the proposed solutions comport with existing international obligations regarding copyright?

Comments and Notes on Comments

More than 700 comments were received, some quite lengthy. You can read them at http://www.copyright. gov/orphan/comments/index.html, or download a 24MB Zip file including all the *distinct* comments. I say "distinct" because 18 comments (if I count right) are identical to an astonishing 126-page illustrators' screed, http://www.copyright.gov/orphan/comments/ OW0660-Holland-Turner.pdf, which in its first four pages denounces Creative Commons, asserts that the whole "orphaned works" concept is part of a movement to "subvert existing copyright protection for other work," and claims that commercial stockhouses would declare huge quantities of materials "orphan" simply to save money. The rest of the massive document is a directory. Apparently listing every member of an organization adds importance to an argument—and having 18 different members send in the same content makes it 18 times as weighty.

The set of responses, far fewer in number, is at http://www.copyright.gov/orphan/comments/reply/.

Most comments note real-world difficulties caused by orphan works—or in some cases what's happened when an organization has decided to take a chance. For example (drawing from a March 10 summary at FreeCulture.org):

- "The DigiBarn Computer Museum is telling the 30-year story of personal computing in an online project with more than 50,000 objects, "many of which are 'orphaned works'…from now defunct firms. We have a statement on every page on our site offering to remove works if the original copyright holder objects. In 3 years we have never once been challenged about any one of the works on our web site… However, as professional and amateur historians we could very much utilize a formal definition of what constitutes an orphaned work."
- "I am a fan of old radio programs of the 1940s and earlier. These programs offer a priceless glimpse into American culture at perhas the greatest time in our history. But for the majority of this material the rights holders, if any, simply cannot be tracked down…" This person argues for shorter copyright terms and an abolition of automatic copyright and concludes: "Material that lies forgotten for decades and suddenly becomes valuable 70 years later doesn't need to generate unexpected profits for heirs or holding companies. That doesn't encourage innovation in any way, which is the real purpose of copyright."
- From a professional artist whose "work depends on good copyright protection": "As an artist, I also know how much powerful synergy arises from the re-interpretation and re-imagining of old works—so making 'orphaned' work available to the public domain more easily is a huge benefit to society which imposes no burden on the absentee owners."
- Several people commented on trying to get old photos—for example, wedding photos *they paid to have taken*—restored and copied by photo shops, to be turned down because the original photographer presumably still held the copyright. (You think you own your wedding photos? Check the contract you signed with the photographer: There's a good chance you only own the prints you purchased, without the right to copy those without paying that photographer an extra fee.)
- The publisher of "a magazine devoted to bringing great illustration art back into the public eye" has run into so many "dead-ends in my copyright searches" that source material is now limited to pre-1923 publications. "Much of our artistic heritage languishes in an uncharted limbo of doubt, and it serves no one and no purpose."
- There's also the other side. "What struck me, because I have worked for many publishers, is that to place a so-called orphaned work in the public domain would be to say 'Go get it,' to the publishers, if the work is of any broad interest." This person, G. Miki Hayden, posits a writer who's "elderly and completely out of circulation," and is disturbed by the idea that a work could be considered orphaned, published as "a minor work" by a publishing company, and seen in print by the family. "Are the family members then to be told that they and the elderly father or mother have no rights in regard to the work?" (Until 1976, the answer was quite clear: *Yes*, since copyright would not have been renewed because the author didn't care.)
- John D. Berry, Native American Studies Librarian at UC Berkeley's Ethnic Studies Library, discussed that library's difficulties in locating copyright holders for newsletters and newspaers generated by Native Americans, Chicanos, and Asian Americans. The library would like to microfilm some of the rare publications, and may be the only library in the U.S. holding the items, but "due to the difficulties locating any participants/owners of aforementioned presses we cannot microfilm or can only do so with great difficulty"—and they can't *share* the microfilm without locating copyright holders.
- One commenter wants to put early (pre-WWII) "World of Knowledge"-style encyclopedias online—but the pictures have, at most, nominal author credits, and the articles may not have authors clearly identified. "The cost and time taken to do thorough research into the identities of the 'creative artists' would prohibit re-publication and therefore would stop the project…until mid-century at least."

Mary Minow discussed the nature of the problem posed by orphan works for libraries, museums and archives (my paraphrase in square brackets):

In many cases, the expertise, time and money is simply not available to even determine whether or not a work is still in copyright…let alone track down unknown copyright owners via probate records and bankruptcy proceedings… I am not aware of any libraries that make use of [the 20-year but very limited] exemption. I believe this is because the exemption is so uncertain. [It] applies only where the copy is made for special listed purposes by the library (not subsequent users), it covers only "published" works not "subject to normal commercial exploitation" and not obtainable, apparently not even as a used copy, at a "reasonable price"; and it insists that the library assure itself through "reasonable investigation" that these conditions have been met. Most importantly, the terms "reasonable investigation" and "subject to normal commercial exploitation" are so open-ended that the risk-averse library does not use the twenty-year exemption. [She goes on to applaud the criteria set forth to define orphan *published* works, in particular the lack of the "normal commercial exploitation" loophole, and the idea of a users' registry for those proposing to use an apparently-orphan work:] Nonprofit users, including most libraries, would only need to file their intent or use, at no cost. Commercial users would follow the same procedure but would additionally need to pay a set fee in a segregated trust account…

Minow also suggests no-cost registration by copyright owners as a requirement to maintain a range of remedies for infringement; this would make it possible for would-be users to contact the owners. But the owners' registry must be mandatory to be useful.

Center for the Study of the Public Domain

This Duke Law School center provided two documents as comments (both published under Creative Commons licenses): The 13-page *Orphan works analysis and proposal* and 8-page *Access to orphan films.*

The first comment states the core purpose of copyright as to "enrich the general public through access to creative works," based on a recent court decision. It goes on to suggest that the *Notice of inquiry* understates the dimensions of the problem:

First, copyright law no longer has formalities… Indeed, many works whose authors do not want copyright protection are now swept, willy-nilly, into the copyright scheme…

Second, the nature of technology means that far more "fixed" works are created than ever before, many thorough non-standard distribution channels, whose record keeping is sporadic at best. Thus the problem will only increase.

Third, repeated retroactive copyright term extensions mean that vast numbers of works whose authors had no reason to order their affairs in the belief that rights will subsist are still potentially under copyright…

Fourth, changes in technology mean that publishing, reproducing, editing and commenting are now potentially within the hands of millions, who could offer restored, edited and revised orphan works to the world on the World Wide Web. Yet at this precise moment, which could be the golden age of copyright, it is probably true that the majority of 20th century culture consists of orphan works. That is certainly true in the case of film… The result is particularly perverse. Having done its job and encouraged initial creation and distribution, copyright now stands as an unnecessary barrier to future dissemination.

Fifth, many modern media simply do not last as long as the copyright term. Without a better scheme for handling orphan works it is likely that we will lose them…forever.

The Center notes that the current system "does little...to benefit the authors" of the occasional "apparently orphan work" actually under copyright management—because would-be users *avoid* such work! "The undiscovered author of an apparently orphan work would actually be better off in many cases with a system that required a reasonable search and notice of intended use, and then gave qualified immunity to future use." After expanding on these points, the Center notes proposed legislation that would require a $1 tax (and registration) fifty years after publication, and every ten years thereafter—a modest proposal, but one that would help in the long run. But, as noted, "It does not address the problem of presently orphaned works."

The last few pages of the comment propose a solution based on seven key principles: clear guidelines, low levels of required search, broad coverage, efficient administration, "notice" for proposed uses that copyright holders can easily search, safe harbor for those who have followed the procedures, and protection of value-added restorers and reusers (so that, if you've added value to an apparently orphan work and the holder turns up, you can continue your use on payment of a specified royalty: otherwise, copyright holders gain unjustly through your added value).

The second comment, *Access to orphan films*, asserts that such films "make up the overwhelming majority of our cinematic heritage"—"a vast treasure trove of newsreels, documentaries, anthropological

films, portraits of minority life in the U.S., instructional films, and even some Hollywood studio productions." Such orphans present a special problem because the works "are *literally disintegrating*"—they're on volatile stock that self-destructs. According to the Library of Congress, *half* of the movies made before 1950—probably hundreds of thousands in all, since one estimate is that 30,000 films had been made by 1917—"are already irretrievably lost." In 1994, LC estimated that 80% of films from the 1920s and 90% from the 1910s had already decayed beyond any hope of restoration. Video doesn't help: "In the words of one expert, 'videotape has much more serious longevity issues [than] Film.'"

Many of the works most in need of digitization for preservation are orphan films, "films of long-term cultural and historical value that are not being protected by commercial interests." In one of the perversities of the Sonny Bono act, the Senate argued that extending copyright would *further* preservation by "providing copyright owners at least 20 years to recoup their investment"—but marketable Hollywood features "constitute a tiny proportion of the surviving film population." For works without identifiable owners, the extension makes matters worse by denying others the ability to restore and display the work.

Most unrestored film isn't in studio vaults; it's mostly in public archives such as LC, the UCLA Film and Television Archive, and the Museum of Modern Art. But those guardians typically don't hold copyright—and a single film can have multiple copyright holders. (The comment then offers a shorter version of the proposed solution in the first Center comment.)

Science Fiction and Fantasy Writers of America
This 14-page comment combines anecdotes and an overall commentary. All but one of the anecdotes express "the difficulty (and in economic terms, impracticality) of tracking down authors of older works in order to reprint them." In the science fiction world, authors and editors are the same people, so the authorial concern with being paid for reprints (aided by the fact that *real* science fiction is not generally a high-paying gig!) is balanced by the desire to anthologize the most interesting little-known stories.

Here's SFWA's summary of the problem—which says a lot about the group's perspective:

> Since works are given copyright protection the moment they are written, there is no ready way to find authors to seek their permission to republish materi-

als, and the penalties for infringement are high, there is a lot of material that cannot be republished because the authors are essentially unlocatable…the cost to locate them, if they can even be located, is often too high to justify the use of the work. Factoring in the 95 years / Life+70 years duration of copyright, a large amount of work is likely to be unrepublishable for over a hundred years and possibly lost altogether.

There have been a number of examples submitted from editors how this has prevented them from keeping important older work in print. Author Spider Robinson noted that much of science fiction's pulp magazine heritage could be lost because by the time copyrights expire, the physical magazine issues may no longer exist. (Some have been archived on microfilm, but not all, and the microfilm copies are of dubious quality.) Examples of losing track of authors after less than a decade were given, demonstrating the likelihood that obscure older works are even more difficult to republish. This includes not just short stories, novels, poetry, etc., but web pages, public newsgroup postings, etc.

SFWA's Orphan Copyright Committee arrived at a "compromise consensus" recommendation. It proposes that the copyright office maintain an official author information directory (making finding authors easier), an explicit methodology be established for proving a work is an orphan and allowing limited use (a reasonable search, public notice for six months, followed by statutory compensation into an escrow fund), publishers be charged a small fee (and some percentage of the escrow should be used) to fund the first two recommendations, and a new remedy of three times baseline statutory compensation be established for cases where work is republished *without* following the "orphan works" procedure, so publishers don't try to slide by. Additionally, "group registration" should be simplified and the procedures for registering freelance contributions to periodicals should be clarified and simplified.

It's refreshing to find an organization that represents the interests of copyright holders seeking balance. SFWA isn't a group of fat cats living high off multimillion-dollar advances; it is a group of people that seems to think things through.

Public Knowledge
This group submitted a nine-page comment. The ten bullet points offered as "suggestions when crafting a [proposed] 'reasonable effort' defense to copyright infringement" cover a wide range and include expla-

nations of why proposed remedies would not violate international copyright law.

Public Knowledge believes an orphan work can be young or old and can be any type of creative work; failed "reasonable effort" searches should establish reasonable, predictable limits on remedies available to copyright owners; and the statute should avoid exact parameters of a "reasonable effort" because that varies from medium to medium and work to work. PK suggests publication of brochures outlining typical parameters, encouraging (but not requiring) sworn statements describing such searches (and having sworn statements serve as *prima facie* evidence in lawsuits), allowing one completed search to serve for other users of the same work, and limiting remedies following an unsuccessful search to $200 per use.

Notes on Comments

Wired News ran a Katie Dean story on April 12, 2005, "Copyright reform to free orphans?" She quotes filmmaker Robert Goodman, working on a documentary of early American picture postcards and frustrated because there's no good way to find the copyright holders. It's a good piece (including some of the anecdotes already noted), ending on a downbeat note. Goodman needs to clear 800 to 1,500 postcards. "I will have to hire a couple of people to do this research. I don't know how long it will take… We will do everything in our power to cross all our t's and dot all the I's, and we're still going to get nailed. Somebody's going to come out of the woodwork… At what point do you eliminate my ability to comment on popular culture because of copyright issues?"

The FAQ in *RLG DigiNews* 9:2 (April 15, 2005) is "Adopting 'orphan works'" (www.rlg.org/en/page.php?Page_ID=20571). Peter Hirtle offers crisp, readable definitions of the problem and why it's larger now. Hirtle summarizes the comments, noting that many of them document problems in using orphan works—while some argue against regulations that might reduce absolute copyright control. None of the comments Hirtle reviewed endorsed the Canadian approach, a cumbersome setup that requires extensive research followed by review by a government board. In commenting on his FAQ, Mary Minow refines his notes to come up with this challenge:

> Can we set forth a bright line determination of "reasonable effort"? If we *don't* have some concrete measure of "reasonable effort," have we really set forth a safe harbor for risk averse folks?

At $150,000 an "incident," Minow's question is a good one.

An April 21 story by Dawn Withers (originally in the Chicago *Tribune*, downloaded from tallahassee.com) includes one of the "damaged photograph" anecdotes and some examples of problems facing libraries and museums—but also includes two fairly astonishing quotes from the "other side":

> "It's being framed as 'rich corporations keeping culture from the public,'" said Brad Holland…artist and co-founder of Illustrator's Partnership of America. "But the public doesn't create culture. Individuals do. This is an effort by a bunch of dedicated and well-funded advocates of free culture."

> "Finding a copyright or not finding one is in the eye of the beholder," said Theodore Feder, president of the Artists Rights Society, which represents the estates of major 20th century artists. "This approach is disastrous from the perspective of copyright protection."

Holland's statement is the first time I'd seen *anyone* suggest that orphan works are caused by "rich corporations"—that's implausible, since *rich* corporations are going concerns, which makes it easy to contact them for licensing. The problems are *failed* corporations—and the individually owned copyrights where individuals have gone missing. The last sentence is consistent with the "slippery slope" in this organization's comment—the free culture bunch is just trying to undermine copyright. The second statement seems to presume that no measure of "reasonable effort" is possible and attacks this as undermining copyright. I find it interesting that the Artists Rights Society represents *estates*, not artists—and if it actually *represents* those estates and has a list of them, then "reasonable effort" would find it.

One document that doesn't identify its creator (but appears to be from the University of California), "Observations on the initial results of the U.S. Copyright Office inquiry on orphan works," includes a systematic 5% sample of responses—taking every 20th one in the order in which they were filed. Thirty of the 34 responses checked said orphan works were a problem warranting a statutory solution; four "believed that the orphan works problem did not warrant any diminution of the existing exclusive rights of authors/copyright holders." Twenty-seven comments are from major universities, libraries, and related organizations; *all* of those comments "acknowledge and present evidence that orphan works are a problem."

Joe Gratz commented on ten of the 140+ *reply* comments in a May 12 posting. RIAA appears to favor a solution to the orphan works problem—but then, federal copyright only began to protect sound recordings in 1972, so RIAA isn't heavily affected. He finds the Illustrators' Partnership reply, like their comment, to be "the most copyright-maximalist end of the spectrum of submissions"—noting that they spend most of their reply attacking Creative Commons and proposals for solutions. "There is a startlingly broad consensus for some sort of action toward a solution to the orphan works problem…"

Finally (for now), Scott Carlson writes, "Whose work is it, anyway?" in the July 29, 2005 *Chronicle of Higher Education*—one of those portions that's free sometimes. Carlson begins with one difficult anecdote: Joseph Siry, a professor of art history, wanting to include a sketch of a building in a scholarly article—and Siry, "who is usually meticulous about clearing copyrights," just couldn't track down rights holders. He used it in his article anyway—but he's uncomfortable about it. "Many scholars, archivists, and librarians have stories like Mr. Siry's." When Carnegie Mellon's library studied a sample of 270 items in its holdings, 22% appeared to be orphan works. When Cornell University librarians tried to clear copyright on 343 monographs for its digital archive of literature on agriculture, it spent $50,000 "and months of staff time"—and still hasn't been able to identify owners of 58% of the monographs. In 47 cases Cornell was denied permission "primarily because the people we contacted were unsure whether they could authorize the reproduction or not," according to Peter B. Hirtle.

Then Carlson quotes people from the Illustrators' Partnership—which argues that registration is an unfair burden (presumably because the illustrations have so little reuse value), but that orphan works procedures "would undermine our ability to control our rights and make a living from the work we produce." IP's Holland *specifically* attacks Lawrence Lessig: "Lessig wants to argue that I need to register everything that I do, or it's an indication that I don't see any commercial value." Or how about this: "Peter Jaszi and Larry Lessig and these characters are all arguing that the purpose of copyright law is to bring work into the public domain as rapidly as possible."

My Take

The line from illustrators and photographers is truly difficult to respond to. They're saying they can't be expected to register their works so that someone can find them in order to license them—that would be an unreasonable burden for hundreds or thousands of items, even if registration was free and online. Which is another way of saying that the works don't have enough commercial value to be worth registering.

But the photographers and illustrators want absolute power to bring the full majesty of copyright law, $150,000 per infringement (plus actual damages), down on anyone who *does* reuse their already-published work after trying and failing to locate them.

It's a classic Gotcha. The groups won't consider a scheme that would mean there was money in escrow if they did find their works had been reused through an orphan works system; after all, that would only be a few hundred bucks, not the Jackpot of copyright infringement.

Most works build on other works. Contemporary copyright law precludes purely derivative works. Orphan works present a range of problems that bedevil libraries, museums, publishers, authors, and others—and enrich almost nobody. It should be possible to craft a solution that retains copyright protections in a manner consistent with the U.S. Constitution while eliminating some of the uncertainty that now exists.

Masthead

URL: cites.boisestate.edu/civ5i10.pdf