

Cites & Insights

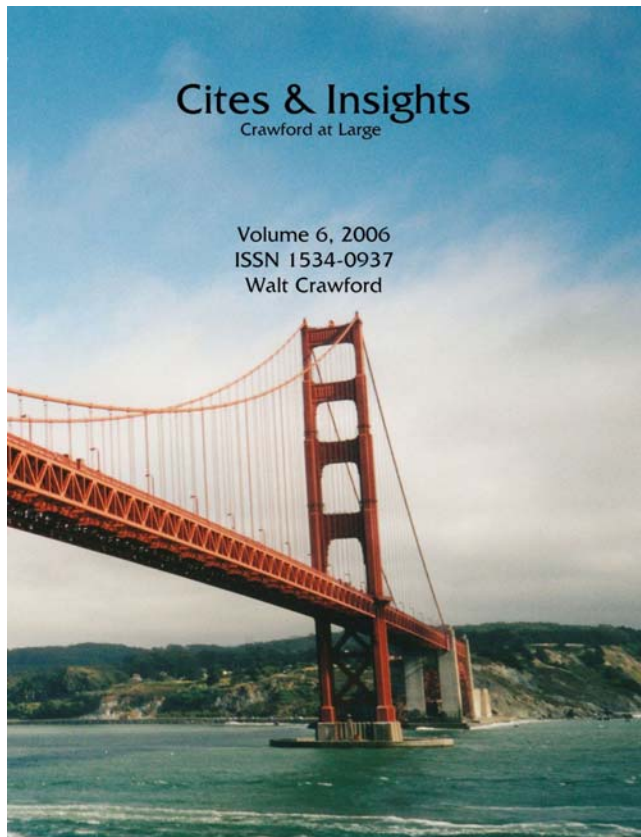
Crawford at Large
Libraries • Policy • Technology • Media

Sponsored by YBP Library Services

Volume 8, Number 1: January 2008

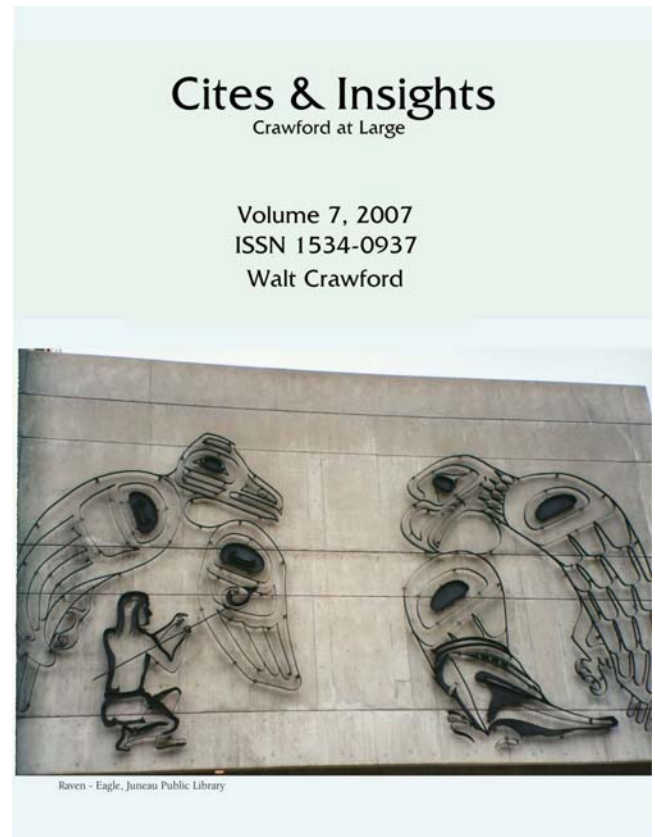
ISSN 1534-0937

Walt Crawford



At one point, a few readers expressed interest in a book version of LIBRARY 2.0 AND “LIBRARY 2.0”, presumably including the followup FINDING A BALANCE: LIBRARIES AND LIBRARIAN. I started on that project but didn’t finish it: Adding the perfect, fully-vetted, Chicago-style citations and bibliography for all of the blog quotes and reindexing the essays seemed like more work than it was worth.

Now you *can* get those essays in book form, indexed and everything—together with another 330+ pages of great content, including LOOKING AT LIBLOGS: THE GREAT MIDDLE, my study of 213 liblogs. For the same price as I would have charged for the Library 2.0 material alone: \$29.50, the (so far) standard price for Cites & Insights Books. Just go to Lulu (lulu.com/waltcrawford) and order it.



Volume 6 had more “regular” pages than Volume 7, but the bound volume’s slightly shorter (388 pages instead of 405 pages), thanks to the extra phantom issue in 2007 (see below).

Inside this Issue

Bibs & Blather: Thinking About Leadership	2
Perspective: Discovering Books:	
An OCA and GBS Retrospective.....	6
Perspective: A Time of Limits?	24

Extra: There’s something extra for buyers of the bound volume, in addition to a great wraparound cover picture (the Golden Gate Bridge from the deck of the Crystal Harmony, passing through on our way to Alaska in, I think, 2001...or 2003...or 2005). A four-page preface offers a few notes about Volume 6,

but mostly has updates on liblogs covered in the 2005 and 2006 studies—which ones have moved and which appear to be gone.

The short version: Of 60 blogs in 2005, 10% may have shut down and 17% have changed URLs and/or names. Of 213 blogs in 2006, 21% may have shut down and another 12% changed URLs and/or names, sometimes more than once (hi, Mermaid!) The preface details all the changes I know about as of December 22, 2007—and clarifies what “may have shut down” means, giving the last post date when it’s even possible to get to the blog.

Don’t Forget 2007

Publication of *C&I* 6: 2006 in book form is the new news—but this is only slightly older, obviously since the final issue of *C&I* 7: 2007. As shown on the right-hand side of page 1, *Cites & Insights* 7: 2007 is also available in book form. Same place (lulu.com/waltcrawford), same price, slightly longer (405 pages), different cover. This time, the front cover is a picture of the sculpture on the front wall of the Juneau Public Library in Juneau, Alaska; the back cover is the sign for Molokai Public Library in Hawaii.

The print bonus for Volume 7? **Cites on a Plane**, the phantom Midwinter 2007 issue that appeared for ten days during January 2007, is included in the book—and not available anywhere else.

These books are great ways to support *Cites & Insights*, and also great ways to keep these essays in a convenient, indexed, easy-to-read form. Both use 60lb. bright white paper (like CreateSpace/Amazon versions of other *Cites & Insights* Books); both retain issue page numbers and issue-oriented indexes.

The combination of Volume 6 and *Balanced Libraries: Thoughts on Continuity and Change* makes a package on Library 2.0-related issues. I’d offer it as a slightly-discounted bundle, but (so far) Lulu doesn’t have provisions for bundles.

Earlier volumes in book form? Maybe, maybe not. Certainly not until the academic library blog book is done. Later volumes? Probably so: The results are so much better than Velobound printouts...

Bibs & Blather

Thinking about Leadership

How could I not? After all, that’s now my primary source of income—not leadership as such, but managing the PALINET Leadership Network (henceforth PLN). For several weeks in November and December

2007, I was editing (a little), marking up (a lot) and reorganizing (wholesale) some three books’ worth of content to get PLN in shape for its 2008 ALA Midwinter formal launch. When you’re editing material—and particularly when you’re adding subheadings and deciding what categories the material belongs in—you tend to *read* the material.

I’ve never been much for reading about leadership. That was 190,000 words ago—or, to put it another way, 72 individual and group essays by library leaders and 15 monthly collections of abstracts and comments on leadership material (and related library material) from a variety of sources.

At times during the process, particularly when assembling the 15 monthly collections and starting to parcel pieces back out into categories, I felt like I was in one of those old Westerns I’ve been watching: **Lots of bullets**, some hitting the mark. Five hints for... Eight guidelines to... Seven ways you should...

It has been and continues to be fascinating. I’m probably picking up lots of useful ideas, most of which would have been more useful (for me) a decade or two ago. I’m *certainly* reading interesting commentary from a range of library leaders—and starting to take steps to inspire more such commentary. The second phase of getting PLN going is probably the toughest for me, and it’s dominated my mental energy to an extent that’s probably evident in this odd issue of *Cites & Insights*. (The first phase was before my time, when PALINET set up the wiki and made arrangements for all that third-party content.) That being the case, I thought it reasonable to offer some comments on the process I’ve been going through—and to add thoughts that may also show up on PLN as an article: “Who’s a leader?”

If you’re a leader, join PLN

If you consider yourself a library leader—present or future—you should be reading and contributing to PLN. It’s not just for PALINET members. That organization is paying the bills, but PLN is open to anyone who reads English and thinks it would be useful.

Point your browser to pln.palinet.org, click on “Log in/create account” in the upper right hand corner and fill out the simple form—a username, password, institution (“retired” or “freelance” or whatever is fine if you’re not employed by an institution), email address and your real name. The real name is optional, but there’s a penalty for not providing one. The email address *must* be real: As soon as you finish entering the account form, PLN sends you email with a link. You need to click on that link to open your account

and be able to read all of PLN. (If you never get the email, you probably mistyped the email address; try again.) It's better if you use an institutional email address, for reasons that should become clear.

The steps above—which take less time to do than to write about—provide read-only *access* to PLN. PLN will work better when people *participate*—commenting on articles, responding to existing articles with new articles or creating new articles because they're worth writing, in some cases adding to or refining articles, participating in forums. However, as most anyone who's dealt with wikis knows, there are two problems with participation: Spam and vandalism.

Thus the second step. Once you click on the link to open your account—or, rather, “a while later” (on a cyclic basis)—another email goes to a special list, one that currently consists of me but will eventually have other PALINET people on it. We (I) look at your email address, institution name, username and real name. If they look “plausible” (institutional email addresses almost always do), we send a reply. That reply authorizes your account to *edit and create* articles, and you get email to that effect. Usually, that step takes place the next business day after you create an account, although it might happen sooner (or, with vacations or problems, later). It's possible that you might get email asking for clarification—but, at least for now, my bias is toward approval: So far (as of this writing in late December 2007), I haven't turned down any account authorization request.

As already noted, there's quite a bit there with more on the way. The first 190,000 words came through a continuing arrangement with Frank Hermes' Library Leadership Network. We've already started adding new content—a PLN Challenge Panel of librarians responding to monthly questions, solicited articles (which may have originated as blog or list posts) and so on. Even before I started putting existing material in shape, I started thinking in terms of new material. Deanna Marcum of the Library of Congress gave an exceptionally interesting luncheon keynote during the 2007 PALINET Annual Conference (my first real encounter with PALINET people). At the end of one section, I jotted down “Great stuff for PLN.” Sent her email a few weeks later asking for permission to use it, which she provided immediately. You'll find it as “Planning for innovation: Experience at the Library of Congress,” along with a link to her full speech. I also found myself reading Jeff Scott's *Gather no dust* with an acquisitive eye—and when he signed up as one of the earliest PLN users, I asked for (and got) permission to use some of his posts. (I have

a small stack of other similar cases and will be watching for more.)

You can't edit everything

As readers of *Walt at random* already know, I raised a question on November 8, 2007 about essays and the “wiki way.” Here's the post; you can go to the blog (walt.lishost.org) for the 20-odd comments.

I have a serious question, particularly for those of you who contribute to or maintain wikis:

Does it violate the “wiki way” for signed content pages—that is, essay contributions with prominent signatures—to be locked against edits (but have open Talk pages)? As a wiki user, would you be offended by such locked pages?

This isn't a hypothetical. I'm working on a fledgling wiki that should become a major resource. It's clear that much of the content will consist of signed essays. Some of those essays will be contributed directly to the wiki. Others will be contributed indirectly (by people who've already written them or are unwilling to deal with wiki markup). Still others will come from third-party sources and those must be locked (as a general rule).

Every locked page will have an open Talk page, open for contributions by anyone with an account on the wiki. We'll try to make the Talk content more visible in a number of ways. When people have substantial alternative viewpoints, we'll link to those content pages from the locked pages.

What do you think?

* As a writer, if you contribute something that should be (and is) signed—anything in the first person, anything with a strong voice, anything that's primarily opinion or your own experience—would you prefer that page to be open to edits by others or would you prefer it locked, edited only by the wiki's editor(s)?

* As a wiki user and contributor, would you be offended by frequent locked pages when they're always accompanied by live Talk pages?

The “fledgling wiki” is the PALINET Leadership Network. I paid attention to all the comments and to an excellent email response I'd received from Meredith Farkas (who I not-so-jokingly call the Queen of Library Wikis) when I asked her the same question. As far as I can tell, most people who *run* wikis in the library field don't see a problem. A few people thought I should blend the wiki with a blog or use a different platform entirely. I also checked a few wikis in the field and found a blend of editable and protected pages to be the rule rather than the exception—although it's clear that some wiki purists regard this as anathema.

We refined the site design somewhat based on the responses. The “wikiness” of PLN has been downplayed—it's still MediaWiki software, but it's really a portal based on wiki software. There's a very clear flag

[protected] near the top of each protected article, along with the author's or editor's signature and publication date. That flag links to a page explaining the policy. Every protected article ends with a "Your turn: Talk about it" subhead linking directly to the **talk** page that accompanies the article. (Thanks to John Houser at PALINET for making the structural changes to help make this work—including changes so the word "talk" is consistently used, not a mix of talk, discuss and discussion.)

Anyway, when you join PLN and start reading, you will find that many items are protected, so you can't edit them—but you can *always* comment on the talk page. If you have a lengthy commentary, write another article, linking from the talk page. Chances are, I'll move that link to the original article's main page as a "Related article" link.

There's the sales pitch. PLN has a lot of good material. More will be coming. We hope it will be a vibrant, active community affair. The formal launch is during 2008 ALA Midwinter, but you can sign up any time. Now, back to the original discussion.

Working on PLN: A Few Notes

I discussed the beginning of this process last time around: BIBS & BLATHER PERSPECTIVE: ON CHARTING NEW COURSES, *Cites & Insights* 7:13 (December 2007). What's happened in the intervening weeks between when I wrote that section and when I'm writing this?

- I looked at existing content and structure and where I thought things *could* go long enough to build a sense of PLN's gestalt—and yes, I think "gestalt" is the best word to describe it.
- Each existing article needed some reworking—to use consistent formatting for contributed essays, to insert the [protected] flag, to clean up minor problems and do light text editing, to rethink category assignments.
- That process helped refine my sense of what the basic set of categories should be. The *full* set of categories will (or *should*) grow over time, but most new topical categories will be subcategories of the basic set. (There are some categories that aren't topical—e.g., categories for recurring features such as the PLN Challenge Panel.)
- I got back to being comfortable editing content I profoundly disagree with, to make it as readable as possible without changing the writer's intent or voice. I was good at that when editing the *LITA Newsletter*. It just requires a little inner voice saying "This isn't *your* platform—you're just the editor." What I can do in such cases is

what I would and will do for others when they have alternative or opposing viewpoints: Add a link to their (or my) alternative viewpoint at the bottom of the existing article. So far, that's happened twice.

- With that process nearly done (for the initial phase—we'll keep importing and marking up content, to be sure), we revised PLN's basic template so a sidebar of "topics" matches the base categories—and, after thinking about it, those topics *link* to category pages rather than separate topical pages. That way, each topic page *automatically* includes a complete up-to-date list of all articles related to the topic (category), in addition to highlighting a handful of recent articles and one or two earlier articles.
- Then, I did something that will send shudders down the spines of experienced wiki folk: After reformatting, recategorizing, editing and moving most articles, and rebuilding the high-level pages pointing to those articles, I deleted several dozen redirect pages that were now superfluous. If there had been significant numbers of users, that would be a terrible thing to do, since article-level URL links would no longer work. That phase is done; I don't anticipate breaking URLs in the future.
- With that phase done, I removed a temporary "Under Construction" banner from the home page, cleaned up the home page as much as possible and began the soft launch of PLN through a post at *Walt at Random* inviting people to join. Several dozen people joined during the first week. Seventy-nine joined between the beginning of the soft launch (December 7, 2007) and December 27, 2007, and they're still coming in slowly.
- The next phase was odd indeed. Leslie Dillon has been contributing excerpts and comments on a wide range of sources under the heading "Leader's Digest." She does this on an irregular basis through a blog. PLN automatically displays the most recent items from the blog using a feed. But older material simply disappears—and a lot of that material deserves to stick around. So I gathered up each month's worth of posts, changed the markup from HTML to wikitext, did light editing for consistency, and saved one cumulative article for each month. Then—the part that's still happening—I copied some chunks to standalone pages, either in related groups or singly, assigning categories to those pages and increasing their visibility.

It's fair to say I'll probably never love wiki markup. It's *certainly* fair to say I'd rather write in Word or Wordpad or Notepad or even the "oops, there goes another

paragraph break” WordPress editor than in the MediaWiki editing box—but, of course, I do much of my direct PLN work within that editor, like it or not.

I’m still learning and expect to continue learning as long as I’m working on PLN. I’m certain it benefits the library community and delighted it’s open to a broader community. The “community” aspect of PLN—direct comments, new articles generated by users, activity in the forums—is just starting and I hope to see it thrive. I’m also acutely aware of the 90:9:1 rule for this kind of application.

If I’m temporarily writing a little less than usual—well, you know, that could be a good thing. I’m certainly reading more (for PLN and on my own), and that’s *definitely* a good thing.

Who’s a Leader?

When you’re managing something that calls itself a “leadership network,” you’d like to know who your audience is. Theoretically, you could *restrict* the audience to “leaders”—but that would require some reasonable way to verify that someone is a leader.

So who’s a leader?

Are managers, directors, other bosses leaders? Maybe, maybe not. Ideally, every manager and director *should* be a leader—but I’m sure many of us have had managers who didn’t qualify. Some of us may have *been* managers who weren’t much at leading. I’m inclined to believe that may have been true of me, back two decades ago. And, indeed, at one point I said, “Well, I can be a neutral party in PLN because I’ve never really been a leader.” That turns out to be nonsense—or at least I hope it’s nonsense.

Are *only* managers, directors, other bosses leaders? Absolutely not.

Consider two of the most obvious cases within organizations:

- **Project leaders** may not have managerial roles as such, but do take the lead for a particular project.
- **Team leaders** may not manage their teams but provide the primary contact point and organizing hub for the team.

And there are other kinds of leadership. For example:

- **Professional leaders**—committee chairs, organizational presidents, all sorts of other leadership roles. I was president of LITA in 1992/93, and I think I was a pretty effective leader at the time. (Similar roles exist in all sorts of other organizations—community, church, charitable, hobby, whatever.)
- **Innovation leaders**—people who bring up good ideas and, more importantly, make them

relevant and workable, bringing them into the real world.

- **Thought leaders**—a tricky phrase, but it applies to many nonfiction writers, some speakers and others who provide leadership for ideas.

That’s a partial list. I’d like to believe I fit in at least two of those three categories.

But wait! There’s more!

This incomplete list doesn’t cover the territory—because it leaves out two groups who are important but also somewhat amorphous: Future leaders and emerging leaders.

PLN directly targets one group of emerging leaders—those in ALA’s Emerging Leaders program. We hope that relationship will be mutually beneficial. Naturally, PLN should also be useful for those who expect to be leaders in the future.

Cue Sly and the Family Stone...

“Everybody is a star” was a great lyric (and a great song—make sure you hear the 1969 original, with three Stones and another band member trading off lead vocal lines, making everybody a star).

Is it a reasonable sentiment in the real world? Can everybody be a leader?

Can, yes. *Will*—probably not: Some people are content to follow.

Should everybody be a star? Should you be a leader in some area at some point? That’s a tougher question. I’m inclined to believe that, at the very least, every professional librarian and a lot of us other library professionals should exhibit leadership in some areas at some points.

Maybe *everybody* in the library field is eligible for PALINET Leadership Network and could benefit from it. It would be nice to think so. As it stands, we’ve opened up PLN to “everybody”—with the caveat that all material in PLN is in English and anonymous or pseudonymous users might not get editing privileges.

A sidebar about pseudonymity: Yes, there are instances in which it makes sense for a PLN essay or comment to be pseudonymous—in fact, there’s already one pseudonymous essay on the site. We’ve made provision for “indirect contributions”—articles that come to PLN via .doc or .rtf or .txt attachments, or just plain email—to pln.content@gmail.com, PLN@palinet.org, or crawford@palinet.org. That’s mostly for the benefit of people who don’t want to learn wiki markup, but if someone has good reason to want an article or comment to appear under a pseudonym and makes that case and request, we’d honor it.

There's a flipside: I don't believe it makes sense for anyone to be a full-time leader in all aspects of life—to lead all the time in everything they do. I think that's hard on the soul, albeit possibly good for the ego. I'm convinced that effective leaders must also be effective followers. Which brings me to the final chunk of this wayward essay:

Leaders need followers

One sure sign that a manager isn't much of a leader is when that manager's employees flee as soon as the chance arises—transferring to other groups or quitting for other jobs.

If nobody follows you willingly, you're not a leader.

You say you're an innovation leader because you espouse shiny new things? Are people taking your suggestions and providing useful new services, solving real-world problems? If so—if you have willing followers—you may be an innovation leader. If not, you're either a voice in the wilderness or a crank.

If nobody follows you willingly, you're not a leader.

I'd almost go so far as to say that leadership is *defined* by willing followers—that the quality of leadership can be directly mapped in the quality of followers. But maybe that's too simplistic.

There's another aspect of this: Leaders aren't necessarily out in front. Just as I was writing this (that is, between writing sessions, during PLN work sessions), a colleague I've never met sent me a note about “leading from behind.” It's an interesting concept, one I hope this colleague will expand into an article.

Making the PALINET Leaders Network a Network
We're off to a great start in providing and organizing thoughtful, interesting, varied content on many aspects of library leadership and leadership in general. But that's just the start.

I'll continue to look for “outside” content that makes sense for PLN, discussing it or importing it as appropriate and possible. We've started the PLN Challenge Panel and hope to see it continue, and to see each set of responses kick off additional discussions and maybe new articles. We'll continue to get some content through other arrangements.

But for PLN to be as effective as it should be, the leaders (of all types, present and future) need to *participate*. Comment on existing articles: There's always a **talk** page, even on essays you can't add to directly. Respond to or expand on existing articles through related new articles. It's easy to write a comment, include “I have more to say in [[Name of my new ar-

title]]” (with the double brackets), save the page, click on the new red “Name of my new article” and proceed to write the article on the page PLN just created for you—or to import and mark up text you've already written. (It's also easy to send me the text and tell me what it responds to: I'm *not* from the government but I *am* here to help.) Poll other users (it's *amazingly* easy to set up a poll!) and respond to polls. Get involved in the forums. Participate.

We'll all benefit.

Perspective

Discovering Books: OCA & GBS Retrospective

Has it been a year since I wrote about Google Book Search and the Open Content Alliance? Apparently so, at least in *Cites & Insights*. Why haven't I written about these projects for a year? Partly because there were too many other things to write about. Partly because the lawsuits don't seem to be either proceeding or going away as the projects continue. Meanwhile, I've been accumulating items worth discussing over the past year, albeit more selectively than in the past. I want to note a few of those items, with comments as usual.

First, a retrospective may be in order. We're now three years into the Google Library Project and just over two years into Open Content Alliance and Google's sensible decision to rename “Google Print” as “Google Book Search.” What follows are extensive excerpts from five major essays on these projects over the past 25 months.

Throughout these looking-back sections, material from previous *Cites & Insights* essays appears as quoted material (indented and in smaller type), with quoted material *within* those essays even more indented (but not even smaller). New comments—primarily updates—appear in regular type. Some material has been reformatted and I don't use ellipses for omitted portions, since the originals are all available.

Where It Began:

Excerpts from December 2005

Much of the first of two December 2005 PERSPECTIVES concerned ebooks and eink/epaper. Most of that's omitted here.

Q&A

Will the Online Content Alliance make ebooks freely available? OCA *does* plan to provide digital facsimiles of

book pages, which taken together constitute one definition of an ebook (not just the etext for a book). That's why PDF will be at least one standard form of OCA availability: It's one way to preserve the design of a printed book. It's also likely to offer PoD.

Update: The Open Content Alliance website seems to be mostly static. Meanwhile, the Internet Archive's Open Library has two sites: A primary site that mostly offers a few "flipbooks" and an "Open Library Demo" site (demo.openlibrary.org) that *may* yield a two-years-in sense of what OCA will offer. Books and articles at the demo site are available in a fairly snazzy page-turning online-reading view, but also as PDFs, in DjVu form, as plain text (based on OCR from the scans) and, potentially, for purchase from Lulu. The current example of Lulu availability raises questions: It's priced at \$12, \$4 more than the production cost for a relatively brief (173 page) public-domain paperback. The demo site is a work in progress. I was unable to find any link from the OCA site to Open Library. Some (most?) OCA books *are* available through the Internet Archive; see later for some counts.

Will the Google Library Project (GLP) make ebooks freely available? GLP allows on-screen reading of digital replicas of book pages, but does *not* allow coherent downloading of complete books. It takes a broad definition of "ebook" to include what GLP provides—but that could change. Karen Coyle has suggested that GLP is "creating a lot of automated concordances to print books," and that's partly true—except that the concordances are bundled into one huge metaconcordance, and for copyright books GLP only shows the first three occurrences of a word or word combination, unlike a proper concordance.

Update: This has changed. Google Book Search now provides single-download PDFs for public domain books and offers a plain text (page-by-page) view along with the page-by-page digital replica. The page-by-page view isn't as snazzy as Open Library but there's a decent "about this book" compilation of metadata and related items—and links to sources for purchase and Worldcat.org for library copies.

Is the in-copyright portion of GLP fair use? In my opinion, it should be—even though I've also said in the past that it probably isn't. Not because Google will be "making in-copyright books available online"—the project is quite clear about *not* doing that, and I can't for the life of me turn three paragraphs of a book into a portion that would violate *any* definition of fair use. The problem is the complete cache that lies behind the full-text indexing and provision of those three snippets: That's a copy by most current definitions and some authors and publishers claim it's copyright infringement. I'd like to believe I'm wrong in my earlier opinion, and lots of people who

know more about copyright than I do seem convinced that it is fair use. The problem with a court trial is that it could either expand the explicit realm of fair use (ideally shifting owner's control toward digital *distribution*, eliminating cached copies as potential infringements), or it could help undermine digital fair use by finding for the publishers and authors. On balance, I hope the court case goes forward—but I'll be surprised if it does.

The court cases continue: **Color me surprised.**

Does GLP harm book sales? GLP will *not* make in-copyright books available for free, and as currently described won't make it easy to read most public-domain books for free. By encouraging discovery for relatively obscure works, Google Print should *increase* book sales, giving a little more visibility to non-bestsellers (the "long tail" if you need *Wired*-inspired jargon for longstanding phenomena).

Studies now indicate that Google Book Search increases book sales, as you'd expect—and it has indeed increased visibility for relatively obscure works.

Does GLP harm authors? How could it harm authors to make their works more visible? Well, OK, it might harm *some* authors—those whose writing or thinking is so bad that three paragraphs turn off potential buyers and those whose works are clearly inferior to lesser-known books that GLP makes visible. The claim that GLP hurts authors or publishers because it deprives them of some theoretical market for making their books full-text indexed online or leasing the books so someone else can do it is, I believe, implausible.

Will OCA and GLP replace online catalogs? I believe the visibility of the first chunk of Google Book Search is starting to clarify this situation. Full-text searching of book-length text just isn't the same as good cataloging, quite apart from the fact that OCA and Google Book Search won't usually provide instant access to local availability or combine circulation with cataloging data. Not that full-text book searching isn't valuable; it is, but its role is complementary to that of online catalogs. The projects might hasten the improvement of bad OPACs; that's not a bad thing.

Will OCA and GLP weaken libraries? I believe OCA and Google Book Search (formerly Google Print) will both strengthen libraries by making works more visible, particularly with links to library catalogs and metacatalogs for local holdings. Even with full-download capabilities, most users are likely to prefer a print copy for those texts that they wish to read at length. Forward-looking libraries will be working to provide links between OCA, Google Book Search and their own services; some already are.

Will OCA and GLP strengthen the commons? OCA should definitely strengthen the commons by making substantial quantities of public-domain material *available*—and, as currently planned, by helping to define the public domain itself by identifying post-1923 books with lapsed copyright. As for GLP, it really depends on how the

project progresses and the extent to which Google decides to cooperate and interoperate with OCA, Project Gutenberg, and other digitization and etext projects. At the very least, GLP will make pages from public domain works available, which strengthens the commons (although not as much as the open approach of OCA).

All four answers are valid today, although OCA now has full cataloging (I think) and GBS has decent metadata. Also, GBS is making PDF downloads of public-domain books available. Open Library makes a bid to be a “universal catalog.” That’s an entirely different set of issues. Unfortunately, there are still no signs of cooperation between OCA and GBS.

Should librarians struggle to assure that OCA, GLP, and related efforts don’t overlap? Chances are GLP will digitize the same “book” (that is, same edition of a given title) more than once if it succeeds in its overall plan. Since OCA isn’t one digitizing plan but an umbrella for a range of related initiatives, it’s even more likely that the same edition will be scanned more than once, particularly when you combine OCA, GLP and other projects. If the digitization really is non-destructive, fast, and cheap, that may not matter. The costs (in time and money) of attempting to coordinate all such projects in order to prevent redundant scanning may be higher than the costs of redundant scanning and storage. As for semi-redundant scanning—that is, scanning more than one edition of a title or more than one manifestation of a work—it’s not at all clear that avoiding such semi-redundancy is desirable, even if feasible. Lightweight methods aren’t necessarily the most desirable for every project; for a loose network of low-cost book digitization projects, however, keeping the bureaucratic overhead light may be essential.

That umbrella turns out to be looser than I thought. As Yahoo! receded into the background, Microsoft initiated its own scanning projects while also part of OCA. I don’t yet see how that’s going to work out.

Open Content Alliance (December 2005)

The best description I’ve seen of OCA is embedded within the FAQ (www.opencontentalliance.org/faq.html). Here’s quite a bit of it, leaving out most questions, with a couple of comments interjected:

The Open Content Alliance (OCA) represents the collaborative efforts of a group of cultural, technology, nonprofit, and governmental organizations from around the world that will help build a permanent archive of multilingual digitized text and multimedia content. The OCA was conceived by the Internet Archive and Yahoo! in early 2005 as a way to offer broad, public access to a rich panorama of world culture.

The OCA archive will contain globally sourced digital collections, including multimedia content, representing the creative output of humankind.

All content in the OCA archive will be available through the [OCA] website. In addition, Yahoo! will index all content stored by the OCA to make it available to the broadest set of Internet users. Finally, the OCA supports efforts by others to create and offer tools such as finding aids, catalogs, and indexes that will enhance the usability of the materials in the archive.

Worth noting: Yahoo! does *not* plan to be the sole source for web searching.

Contributors to the OCA include individuals or institutions who donate collections, services, facilities, tools, or funding to the OCA... The OCA will continue to solicit the participation of organizations from around the world.

The OCA will encourage the greatest possible degree of access to and reuse of collections in the archive, while respecting the rights of content owners and contributors. Generally, textual material will be free to read, and in most cases, available for saving or printing using formats such as PDF. Contributors to the OCA will determine the appropriate level of access to their content...

“Formats such as PDF” is *not* the same as “only available in PDF”

Metadata for all content in the OCA will be freely exposed to the public through formats such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and RSS.

The OCA is committed to respecting the copyrights of content owners...

Will copyrighted content be digitized or placed in the OCA archive without explicit permission from rights-holders?

No... [explained at some length]

The OCA is committed to working with all types of content providers to grow its archive. The OCA has been in discussions with major publishers and the organizations that represent them in order to explore legal, sustainable business models through which more copyrighted content can be made widely available.

There’s the starting point: Something a *little* like GLP—but a lot different, with a broader range of partners, a commitment to openness (including open access where feasible) and interoperability, a strong archival bent, and—on the downside—no single massive source of funding. As Eli Edwards put it in the first blog post I encountered regarding OCA (at Confessions of a mad librarian, edwards.orcas.net/~misseli/ blog/, October 2, 2005): “It is not as ambitious as the Google Print project, but it has the potential to be a very useful supplement, as well as a way to promote open standards and collaboration.

I added this in a January 2006 followup: “By October 31 (2005), OCA had added *dozens* of new members, including libraries such as those at Columbia, Johns Hopkins, Virginia, and Pittsburgh, as well as Smithso-

nian Institution Libraries and others. As reported by Barbara Quint in *Information Today*, there's also some detail on the scanning process. The Scribe system used by the Internet Archive for OCA scanning involves a book cradle with a spine-friendly 90° angle, a glass platen to hold the page flat, manual page turning, and full-color scanning at 'about 500 pixels per inch.' Digitized collections are triply replicated in overseas locations as safeguards." As of December 2007, more than seventy libraries are involved in OCA.

"In challenge to Google, Yahoo will scan books."

That's the headline on an October 3 *New York Times* story by Katie Hafner. The competitive thrust may be necessary to make a newspaper story exciting, but I find it a bit unfortunate. OCA is and should be less and more than a "challenge to Google." Within the story, that's clear: books will be "accessible to any search engine, including Google's." An odd way to mount a challenge! Brewster Kahle snipes at Google, if indirectly: "Other projects talk about snippets. We don't talk about snippets. We talk about books." So, to be sure, does the Google Library Project for *public domain* books, the only kind OCA currently plans to scan. By the end of the article, Kahle's changed from sniping to recruiting: "The thing I want to have happen out of all this is have Google join in."

Kahle went back to sniping later on—among other things, using the incredibly misleading language that Google is "privatizing" public domain material through the Google Library Project.

Microsoft (via MSN) joined OCA in late October. According to an October 25 press release, MSN Search plans to launch MSN Book Search—and MSN committed to digitizing 150,000 books in 2006 (or Microsoft contributed \$5 million for 2006 digitization, roughly the same thing). An October 25 story at *Search Engine Watch* says Microsoft is making separate deals with libraries and will contribute *some* scanned material to the OCA database.

An October 26 Reuters story says OCA has added "more than a dozen major libraries in North America, Britain and Europe." Lulu (called a "publisher of out-of-print books," but I think of Lulu as a publish-on-demand service) is also working with OCA. Google and OCA are talking and it's probably only a matter of time before they find common ground.

Maybe a *long* time...

Just to add a little heat to the light, Tim O'Reilly (whose O'Reilly Media is an early OCA member) grumped on his blog about Microsoft—saying the group was "being hijacked by Microsoft as a way of undermining Google" (according to an October 31 *Seattle Post-Intelligencer* story). When interviewed, O'Reilly backed down, saying "hijacking" was a little strong and that "it's good that Microsoft is participating in the group."

Still, he said he considers it inaccurate to portray Google as the "bad guy" for its initiative and Microsoft as the "good guy" for joining the alliance. In reality, O'Reilly said, the fundamental aims of the alliance and Google aren't opposed.

I haven't seen many commentaries (other than those from AAP and other litigants) calling Google a "bad guy" or Microsoft a "good guy" in this case. Rick Prelinger of OCA and the Internet Archive said, "From the beginning, there was a hope that (Microsoft) would join" and said of its \$5 million: "That doesn't seem like undermining to me."

A November 5 *Washington Post* story notes a Microsoft deal with the British Library that appears to be OCA-related: Scanning 100,000 books (25 million pages) and making them available on Microsoft's book search service next year. "Microsoft says that it will seek permission from publishers before scanning any books protected by copyright."

Google Book Search: The Authors Guild suit

This suit was filed September 20 (2005) as a class action suit with jury trial demanded. The complaint runs 14 double-spaced pages. It claims Google's reproduction "has infringed, and continues to infringe, the electronic rights of the copyright holders..." In the next paragraph, the suit makes a questionable factual assertion:

4. Google has announced plans to reproduce the Works for use on its website in order to attract visitors to its web site and generate advertising revenue thereby.

Google has explicitly said that only *snippets* of in-copyright books, no more than three of them, each containing no more than a paragraph, would be displayed. Calling up to three paragraphs of a book "reproduc[ing] the Work" is outlandish and appears to deny the existence of fair use. The same claim is repeated in the following paragraph.

After a page of claims, the suit identifies three named plaintiffs (Herbert Mitgang, Betty Miles and Daniel Hoffman), each of whom has at least one book with registered copyright held at the University of Michigan (presumably chosen because it's one of two Google 5 libraries that has agreed to full digitization). It then describes the Authors Guild and Google and asserts a class definition and allegations. Paragraph 34 is worth quoting in full:

34. Google's acts have caused, and unless restrained, will continue to cause damages and irreparable injury to the Named Plaintiffs and the Class through:
- continued copyright infringement of the Works and/or the effectuation of new and further infringements;
 - depreciation in the value and ability to license and sell their Works;
 - lost profits and/or opportunities; and
 - damage to their goodwill and reputation.

I'm no lawyer, but it's hard to imagine how points b, c, and d could be demonstrated without showing that Google planned to show a lot more than three snippets from a copyright work—or inventing a new “licensing for indexing” revenue stream that authors never had in the past.

Is GLP fair use?

Even intellectual property lawyers can change sides. William Patry initially called the project “fantastic” but could “see no way for it to be considered fair use... what they have done so far is, in my opinion, already infringing.” He revisits the situation later, analyzing based on market impact, and concludes that GLP is fair use. Jonathan Zittrain (Harvard Law) thinks it's a tossup (or at least that the outcome of a trial will be a tossup).

Yes—or probably

Timothy B. Lee of the Cato Institute says GLP has a strong case based on transformative use and the nearly-certain *positive* market impact. William Fisher (Harvard Law) and Jessica Litman (Wayne State Law) agree. Julie Hilden says yes based on market share but offers a note that seems to confuse justice and law: “But the point of copyright law isn't to protect against *copying*, it's to protect against *harm to the value of intellectual property*.” (According to the Constitution, it's to promote progress in science and useful arts, but never mind.)

Susan Crawford offers a multipoint discussion and says:

All computers do is copy. Copyright law has this idea of strict liability—no matter what your intent is, if you make a copy without authorization, you're an infringer. So computers are natural-born automatic infringers. Copyright law and computers are always running into conflict—we really need to rewrite copyright law. But even without rewriting copyright law, what Google plans to do is lawful.

She uses fair use as the basis for that claim. Her first sentence is unfortunate. As anyone who's ever used a spreadsheet or database, edited a photograph, spell-checked, or used Word stylesheets should know, computers do a whole lot more than copy—but it's true that most of what they do *involves* copying.

Tim Lee has a charming article (**well worth reading**) at *Reason*, “What's so eminent about public domain?” He notes the efforts of copyright extremists to take advantage of the backlash against the Kelo decision (a recent eminent domain case). You get a newly-formed “Property Rights Alliance” talking about “recent Supreme Court decisions gutting physical and intellectual property rights”—but, as Lee says, “there haven't been any recent Supreme Court decisions ‘gutting’ intellectual property rights.” Apparently Authors Guild spokespeople are claiming that GLP “seizes private property” and making an analogy with eminent domain. Lee's note:

Yet in reality, the excerpts of copyrighted books shown by the service would be far too short to be of use to anyone looking for a free copy. And under copyright

law, the use of short excerpts has traditionally qualified as fair use. If the Authors' Guild prevails, it will leave copyright owners with much greater control over how their content is used than they have traditionally enjoyed in the pre-Internet world.

Jonathan Band, who “represents Internet companies and library associations with respect to intellectual property matters in Washington, D.C.,” prepared what may be the most widely-referenced copyright analysis of GLP, “The Google Print Library Project: A copyright analysis.” One version appears in *E-Commerce Law & Policy* 7:8 (August 2005); that version also appears in *ARL Bi-monthly Report* 242 (October 2005), www.arl.org/newsltr/242/google.html. A related article with a different title (“The Authors Guild v. the Google Print Library Project”) appears at *LLRX.com* (www.llrx.com/features/googleprint.htm), published October 15. His concise analysis is clearly written and **well worth reading** in its entirety.

Band notes the need to consider *exactly* what Google intends to do in each aspect of Google Book Search. As regards AAP's attack on Google (and the Authors Guild suit), Band asserts that both the full-text copy and the snippets shown in response to queries fall within fair use. Certainly the market effect seems to favor GLP. Does any rational author or publisher *really* believe that increased findability will decrease their market? “It is hard to imagine how the Library Project could actually harm the market for certain books, given the limited amount of text a user will be able to view.” Band also concludes that GLP is “similar to the everyday activities of Internet search engines” and explains the fair use analogies. Concluding (the *LLRX* version):

The Google Print Library Project will make it easier than ever before for users to locate the wealth of information buried in books. By limiting the search results to a few sentences before and after the search term, the program will not diminish demand for books. To the contrary, it will often increase demand for copyrighted works by helping users identify them. Publishers and authors should embrace the Print Library Project rather than reject it.

Here's how Fred von Lohmann (EFF) sees Google's case for the four elements of fair use as it applies to the Authors Guild suit:

Nature of the Use: Favors Google. Although Google's use is commercial, it is highly transformative. Google is effectively scanning the books and turning them into the world's most advanced card catalog.

Nature of the Works: Favors Neither Side. The books will be a mix of creative and factual, comprised of published works.

Amount and Substantiality of the Portion Used: Favors Google. Google appears to be copying only as much as necessary (if you are enabling full-text searching, you need the full text), and only tiny snippets are made publicly accessible.

Effect of the Use on the Market: Favors Google. It is easy to see how Google Print can stimulate demand for books that otherwise would lay undiscovered in library stacks. On the other hand, it is hard to imagine how it could hurt the market for the books—getting a couple sentences surrounding a search term is unlikely to serve as a replacement for the book. Copyright owners may argue that they would prefer Google and other search engines pay them for the privilege of creating a search mechanism for their books. In other words, you've hurt my "licensing market" because I could have charged you. Let's hope the court recognizes that for the circular reasoning it is. \

I believe von Lohmann's off base on the second point: biographies and other "factual" works are *also* protected by copyright unless they're *purely* listings of facts. As a library person, I could do without "world's most advanced card catalog." Quite apart from being a bit like the world's best jet-powered buggy whip (how many card catalogs have you seen lately?), that description asserts that full-text search is inherently more "advanced" than cataloging, an assertion I disagree with. It's different and complementary.

No—or probably not

David Donahue cites the Texaco case (*no* fair-use right for a private corporation to photocopy entire articles for its research staff) and Williams & Wilkins (fair-use right for a *nonprofit* library to do similar photocopying) and thinks Google falls in between. Eric Goldman says his heart finds GLP "great and therefore we should interpret copyright law in a way to permit it. Unfortunately, my head says that this is highly suspicious under most readings of copyright law."

Karen Christensen of Berkshire Publishing Group doesn't like GLP—and includes an odd attack on Berkshire's primary customer base, librarians:

Librarians, unfortunately, don't understand the rights of the creators and producers of books. Most librarians do not understand the work and expense, the expertise and talent, involved in creating the publications they buy. And quite a few believe that information should be free...

Seeing that comment again two years later, I'm astonished at the contempt some publishers hold for librarians.

Pat Schroeder and Bob Barr go beyond saying GLP isn't fair use. "Not only is Google trying to rewrite copyright law, it is also crushing creativity.... Google's position essentially amounts to a license to steal..."

A November 3 post by Preston Gralla at *Networking pipeline* titled "Google retreats in book scanning project" refers to Google's "plan to make available for free countless thousands of copyrighted books without the copyright holders' permissions." He notes Google is now "not showing the contents of copyrighted books." But that's

not a retreat; it's been Google's *consistent* plan to show snippets of copyright works unless publishers explicitly agree to allow pages to be displayed. He claims the Authors Guild and AAP suits are "no doubt...why no copyrighted books have been made available" and expresses his clear belief that Google should give up: "Here's hoping that Google is having second thoughts about the program, and will ultimately back down..."

ALPSP issued a formal statement stating its firm belief that "in cases where the works digitized are still in copyright, the law does not permit making a complete digital copy for [Google's] purposes." The group opposed Google's opt-out solution and advises its members "that if they are not sure about the program, they should exclude all their works for the time being." On the other hand, ALPSP *does* suggest publishers "protect both in- and out-of-copyright print and electronic works by placing them in the Google Print for Publishers program instead." One wonders how publishers protect out-of-copyright works; surely public domain means public domain? Peter Suber notes that this and an earlier ALPSP statement assert "an abstract property right without claiming injury." The second statement also threatens legal action. His note:

If the ALPSP believes that the absence of publisher injury and the possibility of publisher gain needn't be mentioned because they are irrelevant to its case, then it is mistaken. Apart from their relevance to policy, they will be relevant to any court asked to decide whether the Google copying constitutes fair use under U.S. copyright law.

Should authors and publishers be suing Google?

Xeni Jardin says No, in a September 25 article in the *Los Angeles Times*: "You authors are saps to resist Googling." Jardin's another one who calls the outcome of GLP a "digital card catalog" (do lawyers and writers live in a time warp?). She notes the distinction with the war on file sharing: "Google isn't pirating books. They're giving away previews." Internet history has shown that "any product that is more easily found online can be more easily sold." She notes that the Authors Guild squabbled with Amazon over its "look inside" feature as well. She goes on to suggest that such "paranoid myopia" could lead to a total shutdown of search engines: "What's the difference, after all, between a copyrighted Web page and a copyrighted book?" (Seth Finkelstein's answered that one: Web pages are freely available for anybody to read and download; books aren't.)

Lawrence Solum (University of San Diego Law) also says No and objects to class-action certification: "That class [copyright-holders for books in Michigan's library] includes many authors who would be injured if the plaintiffs were to prevail—including, for example, me!" Solum (a prolific writer) knows he'll benefit from wider dissemination of his works. Jack Balkin (Yale Law) feels the same way: "As an author who is always trying to get people interested in my books...I have to agree...the Author's

Guild suit against Google is counterproductive and just plain silly.” Peter Suber also notes that he’s one who falls into the class and doesn’t want to be included—*unless* Google prefers to fight a class-action suit. Put me in the same category: Michigan owns several of my books to which I hold copyright, and I believe a successful suit will harm me indirectly if not directly.

While Peter Suber admits to seeing plausible cases that GLP infringes copyright, “I haven’t yet seen a plausible case that the authors or publishers will be injured.” He believes the Authors Guild may be looking for a cut of Google’s ad revenue: “If so, then...we’re watching a shakedown.”

Nick Taylor (president of the Authors Guild) thinks it’s necessary to sue—because otherwise Google is getting rich at the expense of authors. He goes on, “It’s been tradition in this country to believe in property rights. When did we decide that socialism was the way to run the Internet?” Going from talking about Google having cofounders “ranking among the 20 richest people in the world” to cries of “socialism” for a Google project: Now *there’s* a creative leap that marks a true Author, as opposed to a hack like me. He brings in “people who cry that information wants to be free” a bit later. Peter Suber comments that Taylor is “as clueless as I feared”—and that’s a fair comment.

You know where Fred von Lohmann stands—but maybe not his analysis of economic harm.

[W]ith the Google Print situation, it’s a completely one-sided debate. Google is right, and the publishers have no argument. What’s their argument that this harms the value of their books? They don’t have one. Google helps you find books, and if you want to read it, you have to buy the book. So how can that hurt them? (From a November 9 *Salon* article by Farhad Manjoo, which Peter Suber calls “the most detailed and careful article I’ve seen on the controversy over Google Library.”)

Playing devil’s advocate: To the extent that Google shows library links as well as purchase links for GLP books, it encourages use of libraries—which some publishers could see as harming sales. But boy, is that a stretch...unless they’re planning to attack the First Sale doctrine next.

Should Google settle?

No: Timothy Lee (Cato Institute): “Given the tremendous benefit Google Print would bring to library users everywhere, Google should stick to its guns. The rest of us should demand that publishers not stand in the way.” Michael Madison looks to Google as a “public domain proxy” and thinks Google should fight the case—even though “I’m not convinced that Google is in the right.” He makes a good point: If nobody ever litigates fair-use cases, what’s left of fair use?

The group weblog *Conglomerate*:

Should Google fight the case? Absolutely. From a litigator’s and trial lawyer’s point of view, this is a case worth fighting... It isn’t very often when a fair use

argument gets raised by a big-time, well-financed corporate entity.

Derek Slater says, “When I look at the Google Print case, I say ‘game on’—I see a chance for a legitimate defendant to take a real shot at making some good law. There’s broad and even unexpected support for what Google’s doing.”

Lawrence Lessig hopes Google doesn’t settle:

A rich and rational (and publicly traded) company may be tempted to compromise—to pay for the “right” that it and others should get for free, just to avoid the insane cost of defending that right. Such a company is driven to do what’s best for its shareholders. But if Google gives in, the loss to the Internet will be far more than the amount it will pay publishers. It will be a bad compromise for everyone working to make the Internet more useful—and for everyone who will ultimately use it.

Yes: Siva Vaidhyanathan thinks it’s the wrong fight: “It’s not just Google betting the company. It’s Google gambling with all of our rights under copyright—both as copyright producers and users.”

Peter Suber notes that the merits of GLP’s case for fair use are important to settle. “But I admit that I’m not very comfortable having any important copyright question settled in today’s legal climate of piracy hysteria and maximalist protection.” He notes that Google’s wealth is a wildcard: It enables Google to defend itself—but it makes Google an extremely attractive target for a class action suit.

The second suit

The American Association of Publishers (AAP) announced this suit on October 19 (2005). While the suit has five plaintiffs (McGraw-Hill, Pearson Education, Penguin, Simon & Schuster and John Wiley & Sons), it’s “coordinated and funded by AAP.” Pat Schroeder’s take in the press release announcing the suit: “[T]he bottom line is that under its current plan Google is seeking to make millions of dollars by freeloading on the talent and property of authors and publishers.”

The suit itself is 14 double-spaced pages (plus a cover page), with many more pages listing copyright titles published by the five plaintiffs and known to be held in the University of Michigan Libraries (along with three Google illustrations that undercut some of the claims, since they show the tiny snippets of copyright books).

Publishers bring this action to prevent the continuing, irreparable and imminent harm that Publishers are suffering, will continue to suffer and expect to suffer due to Google’s willful infringement, to further its own commercial purposes, of the exclusive rights of copyright that Publishers enjoy in various books housed in, among others, the collection of the University Library of the University of Michigan in Ann Arbor, Michigan (“Michigan”).

That’s the second paragraph in “Nature of the action.” The fourth paragraph provides AAP’s assertion of

Google's motive: "All of these steps [in GLP] are taken by Google for the purpose of increasing the number of visitors to the google.com website and, in turn, Google's already substantial advertising revenue." But Google doesn't run ads on its home page and says it won't show ads on GLP pages.

Later, we learn that GLP "completely ignores [publishers'] rights," which is simply false (else GLP would show pages from all books) and get this interesting language on GLP: "When Google makes still other digital copies available to the public for *what it touts as research purposes*, it does so in order to increase user traffic to its site, which then enables it to increase the price it charges to advertisers." [Emphasis added] Quite apart from the questionable nature of the last clause, GLP will not make "other digital copies available to the public" (unless AAP seriously claims that the snippets constitute infringement).

One paragraph appears to dismiss fair use and other restrictions on copyright:

It has long been the case that, due to the exclusive rights enjoyed by Publishers under the Copyright Act, both for-profit and non-profit entities provide royalties or other considerations to Publishers in exchange for permission to copy, even in part, Publishers' copyrighted books.

We are informed once again that each copyright work listed in the exhibits is "at imminent risk of being copied in its entirety and made available for search, retrieval and display, without permission"—never mind that you *can't* search a single book by itself or that "display" consists of no more than three paragraphs, each surrounding an occurrence of a word or term. The suit dismisses any analogy with indexing and caching web pages, partly because "books in libraries can be researched in a variety of ways without unauthorized copying. There is, therefore, no 'need,' as Google would have it, to scan copyrighted books." **Read that carefully:** It appears to say that the existence of online catalogs negates any usefulness of full-text indexing.

Once I saw the first claim of irreparable harm, I read the suit carefully for any claim of *actual* economic harm. The closest I see is paragraph 35:

Google's continuing and future infringements are likely to usurp Publishers' present and future business relationships and opportunities for the digital copying, archiving, searching and public display of their works. The Google Library Project, and similar unrestricted and widespread conduct of the sort engaged in by Google, whether by Google or others, will result in a substantially adverse impact on the potential market for Publishers' books.

That's it. AAP is claiming that making book text searchable and showing a sentence or two around searched words, together with information about the book so that an interested party can borrow or purchase it, "will result in a substantially adverse impact on the potential market for Publishers' books." You have to wonder why AAP

members have cooperated in the Google Publishers' Program if enhanced exposure is such a terrible thing?

A notion that made no sense on its face and has been fairly well disproven since—except, of course, that sales for out-of-print books are likely to benefit used booksellers, not publishers.

GLP and Libraries

Richard Leiter gets it right in an October 18 post at *The life of books*, after clarifying the aims of Google Book Search: "[L]ibrarians need to be prepared for a renaissance; free online services like this will mean better access to libraries and greater demand for books. Not only will libraries' collections grow, but our numbers of patrons will too."

Thom Hickey also gets it right in a November 15 post at *Outgoing*, "Impact of Google print." "Here's my prediction: seeing the page images online will result in more requests for the physical object, not less." He thinks there may also be more use of "equivalent items" (that is, other manifestations of the same work) at other libraries. I'd guess that's also likely.

Barbara Fister posed the libraries-and-GLP question in a strikingly different way at *ACRLog* on October 20: "I can't help wondering—if lending libraries were invented today, would publishers lobby to delete the 'first sale' doctrine from copyright law, arguing it enables a harmful form of organized piracy?"

OCA and GLP Redux (January 2006)

Most of this is quoted commentary—and most of it stands the test of time with little further commentary required. Note another publisher attacking libraries (at least partially); note wildly exaggerated claims for potential damages. Were (are) some of these people taking advantage of sloppy journalism and short attention spans, assuming they might gain some public sympathy for positions they knew to be false? I can't say—but it's hard not to be a little skeptical.

Google Book Search

I was surprised to read on *ACRLog* that "The Ethicist" on *All Things Considered* likened Google's opt-out offer to "a burglar requiring you to list the things you don't want stolen." The Ethicist was talking with Tony Sanfilippo, who in a November 28 essay states that the Google Library Project "is being done outside the scope of traditional copyright protection," dismissing the *possibility* that fair use applies. Sanfilippo says the project "may irrevocably hurt the production of knowledge in the future" and has this to say about the contract (which returns a digital copy of the library's scanned books to the library): "Using an unauthorized full copy as a payment is clearly a copyright infringement."

It turns out Sanfilippo's making a different case: His employer, Penn State Press, wants to sell *its own* digital cop-

ies of books to libraries that already own the print copies. If it can't do that, "many new books won't get published," which turns into this clarion cry: "Do we want to chuck the whole commercial model for the production of scholarship?" And, of course, Sanfilippo uses the term "theft" to describe the situation.

I posted a comment on the *ACRLog* post offering a different analogy from that offered by The Ethicist: "I'll make a photocopy of that poster you printed up to sell, borrowing it from someone you sold it to. I'll index that poster online, telling people where they can buy or see a copy—but I won't show a significant portion of the poster to anyone." I care about ethics as much as anyone, and darned if I can find an ethical problem with that proposition.

A surprising voice *in favor* of GLP being fair use: Sally Morris of ALPSP. Morris says Google agreed with ALPSP that "it was absolutely the case that it is not allowed to [digitize in-copyright material from libraries] *in Europe*." Fair use isn't part of European copyright law; "fair dealing" is narrower. So far so good, but Morris went a little further, in a quote which will no doubt endear her to AAP:

The fact Google recognizes they can't do this without permission in Europe gives us a threshold to work out a way for them to get permission. *In America, they have the law on their side*. Here, they accept they don't. [Emphasis added.]

One publishers' association has gone on record, in the person of its CEO, saying fair use *does* apply in this situation: Google has the law on their side. Amazing.

An odd commentary appeared November 28 in *Times Online*: "Help, we've been Googled!" by William Rees-Mogg, "non-executive chairman" of Pickering & Chatto. P&C is an "academic publisher" that primarily publishes collected editions of major authors, edited and indexed, sometimes with original material added. In other words, they're taking public domain text (in some cases) and adding value. Now P&C's "sturdy, early 19-century business model" is "threatened by a giant 21-century business model, the omnivorous Google." You could say that *many* two-century-old business models have required revision or abandonment in the 20th and 21st centuries. But no. Rees-Mogg says this, referring to "books that are still in copyright and will remain so for 70 years or more" (albeit books that consist predominantly of public-domain text, which he doesn't bother to mention):

If Google can scan these books, without the permission of the publisher, and include them in its database, then most libraries will not need to buy them. And if librarians do not buy them, they cannot be published. The whole world of learning will be damaged, and academic publishing will cease to be a viable business.

Set aside the notion that academic publishing as a whole will disappear if P&C has trouble selling edited public domain works and claiming copyright because of the editing and indexing. This statement makes no sense

unless Google is displaying the full text of in-copyright books. Never in the essay does Rees-Mogg state the *clear, publicly available, flatly stated* truth: No more than three tiny snippets of any in-copyright book will be displayed without prior permission from the publisher.

Here's Rees-Mogg's assertion of the purpose of AAP's suit: "The purpose of this application is to force Google to charge for viewing a copyright book, and to share the profit." Interesting. In his closing statement, he says the very "survival of the book" (not just academic publishing, not just collected editions of the work of dead writers) "depends on" Google "accept[ing] the rights in intellectual property." Which, of course, it does; thus the snippets. (Peter Suber has a briefer and probably entirely adequate comment on Rees-Mogg's assertions: "But this is just wrong.")

Susan Crawford reports briefly on a December 14, 2005 panel talking about GBS; she was a participant. The current argument of publishers is that Google's Library Project can't be fair use because it could affect *potential* markets. That's a pretty good way to eliminate fair use entirely, since almost anything *could* be a potential market. Her comment:

The world is sufficiently unpredictable that anything could happen, right? So fair uses that threaten any possible secondary market can't exist, according to the publishers. In effect, they'd like to use copyright law to protect against network effects and first-mover advantages that they can't personally monetize.

The University of Michigan and Stanford University have both issued recent memos on their relationship with Google. In Michigan's case, it's a "Statement on use of digital archives" dated November 14, noting what the library intends to do with the digital copy of its books that it receives back from Google: preserve the copy in a digital archive, a "dark archive" at least initially (that is, not accessible but there for long-term archiving); define use by the nature of the work (respecting copyright); *secure* the archive for long-term use. It could be used for natural disaster recovery (working with copyright owners), access for the disabled, and possibly computer science research on the aggregate full text. The library will *not* reduce acquisitions because of the digital archive, use it as an excuse not to replace worn/damaged works, or use it to provide classroom access to in-print works. In other words, Michigan will respect copyright, just as you'd expect. "Merely because the Library possesses a digital copy of a work does not mean it is entitled to, nor will it, ignore the law and distribute it to people who would ordinarily have access to the hard copy."

Stanford issued "Stanford and Google Book Search statement of support and participation" on December 7, 2005. The memo says why Stanford's participating in the Library Project (in short, "to provide the world's information seekers the means to discover content") and clarifies that for in-copyright books "this project is primarily supportive of the discovery process, not the delivery process." Google has been scanning works from Stanford since

March 2005, starting with federal government collections (inherently public domain). After those are scanned, Stanford will focus its contributions on works published up to 1964 that are believed to be in the public domain (works between 1923 and 1964 for which copyright was not renewed are in the public domain). The memo also makes clear that “Stanford’s uses of any digital works obtained through this project will comply with both the letter and spirit of copyright law.”... The memo goes on to discuss litigation against the Google Library Project, expressing the belief that courts will find Google’s project to be fair use. It’s a substantial discussion; a piece of it deserves direct quotation:

Historically, copyright law has allowed the copying of works without permission where there is no harm to the copyright holder and where the end use will benefit society. Here, there could be nothing objectionable under copyright law if Google were able to hire a legion of researchers to cull through every text in the Stanford University Libraries’ shelves to ascertain each work that includes the term “recombinant DNA.” There could be nothing objectionable with those researchers then sharing the results of their efforts and providing bibliographic information about all works in Stanford’s libraries that include this term. Through the application of well engineered digital technologies, Google can simulate that legion of researchers electronically through algorithms that can return results in seconds...

Since then, Stanford has gone to some lengths to make those unrenewed 1923-1964 publications identifiable, mounting a database toward that end. If this works, it should be an enormous boon to OCA, GBS and (most of all) the public, as it would open millions of orphaned books for which neither the publisher nor the author saw any reason to bother renewing copyright.

Discovering Books: The OCA/GBS Saga Continues (Spring 2006)

Again, much of this consists of excerpts from the 2006 essay with relatively little new commentary—because this story seems to tell itself, and I find the narrative arc compelling.

Open Content Alliance and related projects

What about the Million Book Project? The stated goal of the project was to scan one million books by 2005. That goal was clearly not reached. Notably, 10,532 scanned books from this project were available at the Internet Archive two years ago—and the number has increased to 10,556 as of March 22, 2006, despite Brewster Kahle’s assurance in December 2004 that “tens of thousands” were on the way. According to MBP’s FAQ, some 600,000 books have been scanned (primarily in India),

but these are not all available online—and, indeed, I can’t find any indication of how many *are* online.

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

The Million Book Project *did* reach the million-book goal, recently claiming 1.5 million books—but as of early December 2007, only 10,696 of those book were available on the Internet Archive’s Million Book Project site, barely 140 more than were there 21 months previously. Whatever’s happening, it’s largely invisible in the U.S. (This appears to have a certain *Through the Looking Glass* feel to it: IA seems to equate “Universal Library” and the Million Book Project—but the Universal Library tab on IA’s Text page has 29,296 items, where the Million Book Project page reached directly has only 10,696. Curiouser and curiouser...)

If you scroll down in search results, you get to the *actual* Million Book Project portal, the **Universal Digital Library: Million Book Collection** at www.ulib.org, a site apparently operated by Carnegie Mellon. That site *does* appear to offer access to 1.5 million books, most of them (just under a million) in Chinese. It’s a distinctly international, “non-Anglocentric” collection, as fewer than 400,000 items are in English (and very few are in European languages).

Google Book Search

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

Cory Doctorow thinks publishers “should send fruitbaskets to Google” and explains why in a February 14, 2006 essay at *boing boing*. I disagree with Doctorow on huge chunks of his argument (print books are going away, people get all their info online, yada yada), but he makes excellent points on some of publisher and author complaints against Google, specifically the idea that because Google intends to make money (indirectly) from GBS, authors and publishers should get a cut of the action. “No one comes after carpenters for a slice of bookshelf revenue. Ford doesn’t get money from Nokia every time they

sell a cigarette-lighter phone-charger. The mere fact of making money isn't enough to warrant owing something to the company that made the product you're improving." It's a *long* essay, particularly for *being being*—4,096 words, the equivalent of more than five *C&I* pages.

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

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

A cluster of articles in *ONLINE* is curious. Marydee Ojala begins with a clear commentary on how GBS actually works, at least in its current form—and hopes that searchability improves as it evolves. K. Matthew Dames argues that library organizations should support GBS—but says "the library community's only public comments on Google Book Search come from an ALA president who seems more concerned with the possibility that his copyright could be 'flaunted' than the possibilities that someone could find, use, or buy his work." I don't understand this: *Cites & Insights* is most certainly part of the library community, as are many blogs and periodicals that have had *very* public statements in favor of GBS. Or does Dames only consider statements by officers of library organizations? David Dillard, speaking from a reference librarian's perspective, thinks GBS can be very helpful when looking for books with relatively obscure content, offers some examples, and concludes "revenue brought in by books should invariably increase as more people learn of books containing answers to their information needs." As with other librarians (whose opinions I've read) who have actually looked at GBS and its potential, Dillard expects it to be a good thing both for book publishing and for libraries. Then there's Michael A. Banks and "An author looks at Google Book Search." It's the same-old, same-old. The illustrations show *entirely* books provided through the Google Publisher Project, showing no snippets at all. Banks claims GBS "can actually discourage some users from buying books" because it "displays the very information being sought" in certain kinds of nonfiction books.

"Having seen the information, there's little chance the searcher will buy the books." That might be true, *if* snippets were more than a sentence or two and *if* GBS didn't suppress snippets in reference works. He speaks of "pillaged" books that are "intellectual property with value, created by people who anticipate being paid for the time, effort, and expense that go into them." Great, except for the preface: "[M]any, many readers buy reference, tutorial, and how-to books to get at specific information. Now they can go to Google Book Search and get the information for nothing." Since that's *simply not true*, the rest does not follow.

Book Searching: OCA/GBS Update (January 2007)

Here's a quick summary:

Google continues to scan books at unknown rates and Google Book Search now includes enough of those books that we can see both the uses and limits of GBS. Google is making public-domain books downloadable, if you don't mind PDFs with "Scanned by Google" on every page. GBS now makes Worldcat and other library searching available more often.

The big October Open Content Alliance spectacular didn't happen. The OCA website shows signs of inattention. If there's an OCA site searching scanned books, it's well hidden.

As noted earlier, that's now available in demo form—but you can't get there from the OCA website.

Despite its early public lead, Yahoo! doesn't have *any* visible presence as a source of book-related information or scans. Microsoft has introduced a beta version of Live Search Books, part of the rebranding of MSN Search and based on Microsoft's OCA scans. Those books are *also* available as downloadable PDFs—if you don't mind a "Digitized by Microsoft" watermark on each page. So far, the interface only offers the books themselves, with no "Find in a library" or "Buy this book" links.

The Internet Archive includes 35,000 books scanned as part of OCA (as of early December), including some—but apparently not all—of those at Live Search Books. These are also downloadable as PDFs—the exact same PDFs as on Live Search Books, for those books scanned thanks to Microsoft.

The Internet Archive includes quite a few different text sources. As of December 7, 2007, here's what I see:

- **By scanning agency/sponsor/donor:** 188,019 items scanned or sponsored by Microsoft; 5,085 sponsored by Yahoo!; 888 sponsored by the Sloan Foundation—in addition to those scanned from and in some cases sponsored by libraries.
- **By type of library or source:** American libraries, 127,152 items—and here the University of California seems to be the biggest player, with 114,580 items; Canadian libraries, 78,519

items, with the University of Toronto appearing to account for more than 90% of them; a number of other tabs including the already-noted small portion of the Universal Library, Project Gutenberg, and several other collections that may or may not overlap with the American and Canadian library categories (just as institutions within those categories have clear overlaps).

How many books has OCA scanned to date? I wouldn't hazard a guess, but "in excess of 300,000" seems reasonable as a minimum estimate.

Google Book Search

An April 18, 2006 item at *OptimizationWeek.com* offers notes from John Wilkin's April 3 talk on the University of Michigan and Google, held at Ann Arbor's public library. Wilkin estimated that the UM portion of Google's project, digitizing seven million bound volumes, would be completed by July 2011—and noted that UM had been digitizing books at a rate of 5,000 to 8,000 volumes per year until Google came along.

Google issued a short series of *Google Librarian Newsletters*, the final one appearing in June 2006. That issue included an introduction to GBS by Jen Grant (product marketing manager), noting that founders Page and Brin asked this question early on: "What if every book in the world could be scanned and sorted for relevance by analyzing the number and quality of citations from other books?" Apart from the usual Googlish simplification as to what "relevance" means, it's an interesting way to lead into GBS. Discussing problems inherent in the fact (credited to OCLC) that only 20% of extant books are in the public domain, Grant cites an estimate that only 5% are in print—which seems likely. "That leaves 75 percent or more of the world's book in [a twilight zone]." Given the GBS goal "to build a comprehensive index that enables people to discover all books," Google needed a way to handle the "twilight zone" books—thus the snippet approach.

Ben Bunnell (another Google manager) offers "Find a page from your past" in the same issue, beginning "The idea that within our lifetimes, people everywhere will be able to search all the world's books from their desktops thrills me." Bunnell notes examples of "interesting uses" of GBS for family research; it's an interesting commentary that stresses GBS as a way of *locating* books that might be of interest, not primarily a way of *reading* them.

I contributed "Libraries and Google/Google Book Search: No competition!" to the same issue. I focused on locality, expertise, community, and resources—four "reasons libraries don't need to fear Google Book Search or Google itself." Briefly (since the article's readily available):

* Every good library is a *local* library—and libraries do local better than Google.

* GBS "will be a fine way to discover the more obscure portions of books, and obscure books in general. But librarians and library catalogs offer *expertise*—

professional education and knowledge to guide users whose needs are out of the ordinary, and classification methods to support comprehensive retrieval and guide people to the materials they need."

* "Good libraries aren't just local libraries. They're *places* that serve their communities in that regard. Good libraries build and preserve communities. 'Cybercommunities' can be fascinating—but the physical community continues to be vital." I note that Google can strengthen a library's role in the community.

* "Need I state the obvious? Google Book Search helps people *discover* books. Libraries help them *read* books."

I also took Google to task somewhat—which delayed publication of the article and resulted in a Google response from the editor. My grumps:

* Many Google Book Search books published prior to 1923, *necessarily* in the public domain, show only snippets when they should show the whole book. The same is true for quite a few government publications almost certainly in the public domain within the U.S.

* There should be a "Find this book in a library" link for *every* book that originates in the Google Library Project and for every book in the public domain. That wasn't the case the last time I tried date-limited searching.

* Ideally, *every* result in Google Book Search should include a "Find this book in a library" link—after all, even books supplied by publishers show purchase links for sources other than the publisher. If Google Book Search is to be a great way to discover books, it should include all the great ways to *get* the books.

Summarizing the responses, the editor said Google was digitizing quickly and would change some books from "snippet view" to "full view" later on—and Google agreed on the second and third points. Google Book Search *does* now show either "Find this book in a library" or "Find libraries" on all or almost all book results, and that's a significant improvement.

"Find this book in a library" now seems to appear on essentially all book results—and on book summary pages, it's under the heading "Borrow this book."

The Ubois commentary: In August (2006), UC announced it would join the Google Library Project. One early commentary struck me as extreme: "Google 'Show-times' the UC library system," posted August 13, 2006 by Jeff Ubois at *Television archiving*. Immediately noting that this was a "secret agreement," Ubois presumes the agreement "may enrich Google's shareholders at public expense." After quoting Brewster Kahle about providing "universal access to all human knowledge, within our lifetime," Ubois says "[I]t's troubling to see public institutions *transfer cultural assets*, accumulated with public funds, into private hands without disclosing the terms of the transaction." [Emphasis added.]

How is UC transferring assets? It's lending books, which will be returned (they never leave the building in most cases). That's (part of) what libraries do. As for "without

disclosing,” it doesn’t take much research to find out that California is (like Michigan) a state in which that “secret” contract was only secret until someone filed a formal request to see it, since it involved a public agency. “UC should expect and welcome public comment if its inventory is effectively being privatized”—but that’s **not what’s happening**.

Ubois presumes that Google’s contract must be like Showtime’s offensive contract with the Smithsonian, which did provide exclusive access for some length of time—thus the neoverb in the post title.

UC’s agreement is probably not explicitly exclusive. But as a practical matter, scanning doesn’t happen twice...

This deal will be costly for UC in staff time and other resources, and the chances that another vendor will come through and duplicate the work are slim.

This discussion is based on pure speculation—and happens to be *false*, since UC was already an OCA partner and Microsoft was *already scanning* UC books and documents!

More than 100,000 of them to date via Microsoft and other means, as noted above.

Ubois makes things worse: Assuming Google’s efficient, it won’t scan a Berkeley copy of something it’s scanned at Harvard, and restrictions may make it difficult for Berkeley to borrow Harvard’s digital copy. “The student of 2012 will have a choice: go to the complete digital library, owned by Google, or go to the partial digital library of his or her own university.”

That’s nonsense. The student of 2012 won’t be able to get the book from Google’s so-called digital library anyway if the book’s not in the public domain, which means the student can do exactly what he or she can do now: Go read the actual, honest-to-gods, printed book, either UC’s copy (if there is one) or one loaned from another library.

Then Ubois asks a series of questions, at least some of which make the same assumptions. For example: “Is it reasonable to ask the public to pay a second time...for material already purchased, simply because it’s now necessary to convert the format in which it is stored?” But UC is **not** “converting the format” in which books are stored. It’s adding new search capabilities to find print books, which still exist as print books.

Ubois concludes, “By acquiescing to Google’s demands for secrecy, UC has compromised the public interest, and set a dangerous precedent for the rest of the academic community.” Which is *truly* strange, given that UC is by no means the first academic institution to sign a confidential Google contract, unless we assume that Stanford, Harvard and Oxford aren’t prestigious enough to set precedent. And given that UC (and probably Google) knew the “secret agreement” could not legally be kept secret.

The contract was posted later in August. A *Computerworld* story notes that the contract grants Google sole discretion over use of the scanned material in Google’s services, which is scarcely surprising—and that it explicitly prevents charging end-user fees for searching and

viewing search results or for access to the full text of public domain works. UC also agrees not to charge for services using the scanned material (excluding value-added services) and that it won’t license or sell the digital material provided by Google to a third party, or distribute more than 10% of it to other libraries and educational institutions. Finally, Google promises to return the books in the same condition (or pay for or replace them) and has 15 business days (three weeks) to scan a given book.

Karen Coyle compared Michigan and UC contracts carefully. She notes that UC’s contract is silent about quality control for the scans (probably a good thing, given GLP’s early results)—and that UC managed to get “image coordinates” so they can highlight searched words on displayed pages (not in Michigan’s contract). There’s a lot more to Coyle’s analysis, posted August 29, 2006 at *Coyle’s InFormation*.

Phil Bradley spent some time with GBS and commented in an August 31, 2006 search on his blog, “Google Book Search—to download or not download?” You’ll get the tone from the beginning:

In theory Google Book Search now allows users to download out of copyright books for nothing. In practice, it’s the usual Google botched disaster that we’re getting used to.

Bradley notes that it’s difficult to find books you can download—and when you do, “they’re often either so old [as] to be illegible, or they’ve been badly scanned so it’s almost impossible to read.” Bradley tried some Shakespeare, to compare the results “with the Google disaster that is Google’s Shakespeare Collection.” He found 14 (of 23 searched) that he could immediately download, although “most of the editions would have been difficult to read, to say the very least”—but that’s better than the three at the special collection.

Finding a downloadable book at Google, I noted the special page that comes along. It’s an interesting document and includes usage guidelines, fortunately *after* saying “Public domain books belong to the public and we are merely their custodians.” One interesting guideline: “Maintain attribution”—specifically, don’t remove the Google watermark from each page. That’s not an entirely unreasonable request, and it’s stated as a request, not a demand. There’s another: “Make non-commercial use of the files.” The books themselves are in the public domain, which means you’re perfectly free to make *any* use of them—but Google’s asserting a right in the scanned version.

A September 4, 2006 post by Bill McCoy on his Adobe blog questions Google’s “pseudo-license” and repeats Ubois’ assertion, in a different manner: “Just because you’ve got a huge pile of cash and were first in line with a cozy no-bid deal to do this scanning—a deal that *cannot even be repeated given the wear and tear on collection items*—doesn’t create a special exemption to [public domain].” [Emphasis added.] But Google and OCA both

assert that their scanning methods create no more wear and tear than reading a book. McCoy's assertion doesn't work for books that circulate and certainly doesn't work for UC (as one example). McCoy's counter-examples are flawed. Google is not claiming ownership of public domain works, *only of its scans*. Google isn't preventing libraries from lending the books that Google scanned and anyone (Microsoft, Yahoo, me) is free to scan a borrowed book and, if it's in the public domain, do anything we want with *our* scan.

By October, some publishers were beginning to admit that GBS is *helping* sales, as reported by Jeffrey Goldfarb in an October 6, 2006 Reuters story. Oxford University Press estimates that a million customers have viewed 12,000 OUP titles (from the Google Publisher segment of GBS). Springer Science + Business reports growth in backlist sales based on GBS. Penguin finds more success from Amazon—and specialized publisher Osprey found healthy growth from both sources.

Karen Coyle posts an important lesson from early GBS scanning in an October 24, 2006 post at *Coyle's InFormation*: "Google Book Search is NOT a library backup." GBS uses uncorrected OCR, which "means that there are many errors that remain in the extracted text" (including all line-break hyphenation). Also, it's not digitizing everything: Some books are too delicate, some will be problematic. "Quality control is generally low" (she provides egregious examples). None of this came as a surprise to most digital librarians, according to a comment from Dorothea Salo.

Péter Jacsó reviewed GBS for *Péter's digital reference shelf* (downloaded November 3, 2006); it's an extensive and negative review, **well worth reading**. He notes the "ignorance, illiteracy and innumeracy" of the software—"OR" searches yielding fewer results than one of the two terms (or *more* results than the sum of the two terms!), limits that don't work, inconsistent handling of full-view books, confusing hit counts. Google doesn't say how many books are in GBS (or in the full-view portion), always problematic for a database. There's a lot more here, and although some of it seems based on using GBS as a source for actual reference information rather than a way to find books, it's nonetheless a good, tough review.

OCA and Live Search Books

Sixty-odd people attended an OCA workshop in October 2006—but as of mid-December [2006], the OCA website shows the October 20 event as being in the future. The website for the OCA workshops has a faulty digital certificate; the "discussion area" has eight discussion sections, only one of which has any topics (that topic consisting of one anonymous post with no responses). The "Next Steps" page claims a November 2006 update date but appears to date from late 2005. The FAQ says "All content in the OCA archive will be available through the website. In addition, Yahoo! will index all content stored by the OCA to make it available

to the broadest set of Internet users"—but there's no search function on the OCA site.

Fortunately, while the OCA level seems moribund, there's some action within the ranks—although not, as far as I can tell, by Yahoo!, the partner with the highest initial profile.

Microsoft made good on its October 2005 promise to join OCA and to release a book search service. Books.live.com went live (in beta) on December 6, 2006. A December 6 post at *ResourceShelf* offers an excellent brief history of LSB, including links to earlier stories. Gary Price focuses less on competition than on choices: "The more options and tools information professionals have the better. Even Google's CEO, Eric Schmidt, has said that search is NOT a zero-sum game."

Microsoft plans to integrate book content with the rest of Windows Live Search, presumably with an available limit for books only. The beta release includes "noncopyright" books from UC, Toronto and the British Library, with books from NYPL, Cornell, and the American Museum of Veterinary Medicine coming soon. Price notes some features of LSB and that "Scanning looks nice from what we've seen." (I put "noncopyright" in quotes because LSB includes quite a few oral histories from Bancroft's Regional Oral History project that are much more recent than 1923, and those don't appear to be in the public domain.)

Tom Peters comments on LSB in a December 12, 2006 post at *ALA TechSource*. "After playing around for an hour or so...I have to admit—against some vague sense that my better judgment is failing me—that I like it." Unfortunately, Peters follows that by repeating a report that "LSB does not work well—or at all—when using browsing software other than Internet Explorer." That's generally not the case; most users of other browsers (certainly including Firefox) have used LSB without difficulty. Peters does interesting searches—and offers interesting comments. He doesn't like the name of the service, but that's really an issue with Microsoft's online services in general. He wonders why there's no overall count for the collection—as do I, although the same can be said of GBS and Amazon.

After reading Peters' post, I did a little experimenting using his favorite search terms ("phrenology" and "spontaneous combustion"). Here's what I found:

*LSB yielded 687 items for "phrenology" and was only willing to show the first 250 of them. It yielded 219 for "spontaneous combustion" (as a phrase; Peters' 660 must be the two words, which yield 887 on December 15, 2006), and would show all 219 of those. (There appears to be a firm limit of 250 viewable results in the current LSB, as the 887-book result also stops at 250.)

*Google Book Search yielded 2,618 for "phrenology"—but would show only 139 books, indicating a typically wifty total result count. For the phrase "spontaneous combustion," GBS showed 1,041, of which 512 were actually available.

*Restricting GBS to full-view books reduced the first result to 1,603 and the *actual* result to a mere 63, either one-quarter or one-tenth of LSB's result. The second search came down to 699 claimed, 489 actual.

Let's redo those searches as of December 7, 2007:

- **Live Search/Books:** Phrenology: 3,450, of which 2,670 are viewable—but as usual, the (remarkably annoying) results page stops at 250. “Spontaneous combustion” as a phrase: 1,960, including 1,670 fully viewable—still with the 250-book limit. The viewer works nicely, and PDF downloads are available. Those results represent *enormous* increases from a year ago—assuming they're real. Since there's no way to limit results (that I could find, at least not within the Books page), it's hard to say for sure.
- **Google Book Search:** Phrenology: 2,080, including 2,372 full view and 2,680 limited view: Google's curious result counts strike again! There actually appear to be **211** full-view books; I'm guessing that those books might include 2,372 pages with “phrenology” on them. Mysteriously, there are actually 289 limited-view books—but only 286 books in total, not the 500 you'd expect to see. “Spontaneous combustion”: 1,890 in all, including 1,007 full view (which turn into 385, all viewable). Also substantial increases—and Google presumably sticks with its usual 1,000-record limit and *does* offer a variety of search refinements. Too bad the raw counts make no sense at all.
- **Universal Library (at ulib.org):** This searches only titles or authors. “Phrenology” as a title search yields ten results, but four of the ten have “0 pgs.” No match for the phrase “spontaneous combustion”; one zero-page match for the two words.
- **Internet Archive texts** (including OCA, Universal Library, Project Gutenberg etc.): “Phrenology” yields 25 results, “spontaneous combustion” two results (as words—a phrase search malfunctions). Note that these are *not* full-text searches.
- **Demo Open Library:** “Phrenology” three full-text items (85 total). Spontaneous combustion (words): one full-text (20 total). Both results show limit sidebars very similar to (and quite possibly based on) Worldcat.org—but Worldcat.org itself yields considerably larger results, “about 2,431” for phrenology (including 1,923 books) and “about 490” for the phrase “spontaneous combustion” (including 202 books). Note that full-text searches on the demo Open Library site don't seem to work yet.

I won't even attempt to draw conclusions based on this study—except the usual one, that Google's raw result numbers are slightly worse than meaningless.

Bringing It Forward: Notes from 2007

A few items worth commenting on, mostly in chronological order, combining all projects. I'm skipping most items, both for space and because some arguments are more tedious than others.

In case it hasn't been clear already: Yes, some of Google's scans are sloppy. No, Google didn't negotiate exclusive contracts and several GLP partners are also involved in other mass scanning projects. Yes, Google is much too secretive about what it's doing. Yes, Google has (at least in the past) been far too cautious with regard to the public domain nature of government information. I think there's a *lot* about GBS and GLP that could be done better.

I scrapped much of what I had because the continued paranoia and repeated arguments become stale after a while. At least one professor seems to be making a career of Google-bashing and I suspect he's not alone. Google can fight its own battles; I find the whole situation sad and disappointing. GBS is no substitute for a library (nor does it claim to be); it is a remarkable, if far from perfect, way to find books you didn't know existed. So is Live Search Books. So, eventually, might be Open Library—but in that case, there clearly *are* grandiose claims, nearly as grandiose as Google's founders' utopian visions.

Dead plagiarists society

Paul Collins posted this “culturebox” piece in *Slate* on November 21, 2006. He considers the use of Google—and specifically Google Book Search—to investigate plagiarism: “For any plagiarist living in an age of search engines, waving a loaded book in front of reviewers has become the literary equivalent of suicide by cop.”

His example is a *Washington Post* book review of Amir Aczel's *The Artist and the Mathematician* in which the reviewer (Charles Seife) accused Aczel of lifting text from Guggenheim Museum's website. Aczel wrote an irate letter to the post, saying “It seems that Seife has submitted every sentence in my book to a Google search.” It isn't just new books. A linguist who works for Google ran a phrase from England Howlett's 1899 *Sacrificial Foundations* through GBS (to find the Howlett book) and came up with a “suspiciously similar passage”—and lots more similarities—in Sabine Baring-Gould's 1892 *Strange Survivals*. But then, Baring-

Gould seems to have picked the sentence up from Benjamin Thorpe's 1851 *Northern Mythology*.

As Collins notes, these are mostly "forgotten writers," but this idle discovery could "become a literary earthquake"—what if scholars start doing extensive automated GBS searching for plagiarism? The corpus is already big enough to yield interesting results (and Live Search Books would add more—as would Open Library when it starts working for full-text searches). Collins then asks the obvious question: "Don't people accidentally repeat each other's sentences all the time?"

Collins says, "It seems to me that this should not be unusual"—and then searches that sentence on Google Book Search. Zero results. (The same is true for the quoted sentence that ends the previous paragraph.) Collins did the "It seems" search as a series, starting with the first word (rejected) and adding one word at a time. "After just a few words, the likelihood of the sentence's replication scales down dramatically." And, as he notes, the nine occurrences of the sentence missing its final word are from a "grab bag of sentences"—finding precisely the same sentence in a work on the same topic seems less than likely. I would note that this is probably not the case for descriptive nonfiction sentences, at least taken one at a time: After all, there are only so many ways to state a fact. (That sentence, not including "after all," appears twice in a Google search—both discussions of plagiarism—but not in Google Book Search.)

This is interesting stuff. Will any "deeply idiosyncratic" author (e.g., Emily Dickinson or Ben Franklin) get fingered? It's already happened, using earlier tools, to Lawrence Sterne—who apparently copied a diatribe against plagiarism in *Tristram Shandy* from Robert Burton's *Anatomy of Melancholy*.

Internet Archive claims progress against Google Library initiative

That's the headline on a December 20, 2006 *InformationWeek* story—and you have to wonder about "against." The story is a \$1 million Sloan Foundation grant, along with a claim that IE had already scanned more than 100,000 books at that point, but it's damaged by Brewster Kahle's statement: "Google has made a full-court press toward privatizing every library they can get a hold of. But this is a step toward showing there's an alternative path." Kahle goes on to claim that institutions who've signed up with Google are usually unwilling to take on the added expense of working with another book-scanning group, saying "It's effectively exclusive." Two major things wrong with that statement and reportage: Google's scanning

should not be a significant institutional expense—and the University of California is one enormous demonstration that exclusivity isn't very effective.

I'd forgotten about this piece—and as a result, thought that Kahle's "privatizing" statement in October 2007 was new. At that point, I wrote an angry post in *Walt at random* entitled "misusing the language." Since Kahle started this misuse early on, this may be the time to quote myself:

I'm not in love with Google by any means. I think OCA is a great idea (although I wonder where the "alliance" has gone, given Yahoo's almost-total silence and Microsoft's diverging effort). But "privatizing the library system" or, which I've also read, "privatizing the public domain"—I'm sorry, but hoespucky.

If Google negotiated exclusive contracts, maybe. Otherwise, that language is like saying that, if I check a book out from my library that happens to be in the public domain, scan it, and return it to the library, I've "privatized" the book.

Google is borrowing books from libraries (in large quantities thanks to special arrangements), scanning those books, and returning them to the libraries with the promise that the books won't be damaged. Its deals are nonexclusive. Google's scan does not in any way modify the terms under which the book itself can be used.

Google Book Search absolutely expands findability for books and in no way restricts anyone else from building and maintaining book-search systems. Google Book Search for public domain absolutely expands access to the text within books, and in no way restricts anyone else from providing similar access. (For that matter, Google's silly first-page "conditions" are suggestions for use of their PDFs, not legal restrictions.)

How can expansion be viewed as contraction? How can improved access be regarded as privatization?

Want to attack Google? Fine. But is it necessary to debase the English language to do so? Or does it just make a great soundbite?

Will mass digitization projects need to be re-done?

Jill Hurst-Wahl uses that title for a December 27, 2006 post at *Digitization 101* (hurstassociates.blogspot.com). She quotes a Joseph Esposito list post asserts "four specific requirements" for mass digitization projects such as GLP: an archival approach, reader's editions as well as digital facsimiles, use of a technical environment that enables ongoing annotation and commentary, and file structures and tools that allow machine processing of the content.

The answer to Hurst-Wahl's question is simple enough: Yes, if you want archival conversion, at least for GLP—because that's not what GLP was aiming for. OCA, maybe. Meanwhile, Columbia's Stephen Paul Davis wrote a lengthy comment calling Esposito's

“musts” foolhardy. “The good news about Google-type mass-digitization, in my view, is that almost nothing these projects accomplish will in any way raise the cost of enhancing or redoing their output in the future—in fact just the opposite.” Which is also the first and most obvious response to Brewster Kahle and others when they assert “privatization”: Google’s scanning of a book *in no way* prevents or hinders later scanning. Davis notes that GLP partners are already holding back fragile materials, the one area where there might be a concern. Davis also goes back to Esposito’s original statement and finds some of it condescending as well as beside the point.

An odd controversy over GLP’s public-domain downloads

Danny Sullivan posted “Authorama: Testing if Google can restrict public domain books it offers for download” at *search engine land* (searchengineland.com) on January 10, 2007. The setup is that the PDFs for public domain GLP books start with a special page from Google and include a “Digitized by Google” watermark on each page. Here are the guidelines, which Sullivan calls “this warning and document guidelines”:

This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world’s books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that’s often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book’s long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google’s system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google “watermark” you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can’t offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book’s appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

Warning? The only warnings in that statement are that public domain works differently in different countries and that Google doesn’t care for (and will block) automated queries (but may be able to provide alternative means). Otherwise, there are *requests*—“we request” and “please” are hardly the stuff of warnings. But the second sentence of the post, far above those (to me) innocuous guidelines, is “Can Google dictate that public domain books that it has scanned and distributed on the web really be subject to restrictions on non-commercial work?”

Phillipp Lensen uploaded 100 GBS downloads to another site to allow redistribution or any form of use. Danny Sullivan asked Google “what they think about the project and the legality of trying to impose restrictions on public domain books, just because they’ve scanned them.” And Google responded (in part):

The front matter of our PDF books is not a EULA [end user license agreement]. We make some requests, but we are not trying to legally bind users to those requests. We’ve spent (and will continue to spend) a lot of time and money on Book Search, and we hope users will respect that effort and not use these files in ways that make it harder for us to justify that expense (for example, by setting up the ACME Public Domain PDF Download service that charges users a buck a book and includes malware in the download). Rather than using the front matter to convey legal restrictions, we are attempting to use it to convey what we hope to be the proper netiquette for the use of these files.

In other words: *There’s no story here. Move along.* Comments include a number of interesting points.

Google probably *could* include a EULA before the PDF download and contractually oblige users to restricted use (in consideration of receiving a free copy)—but they chose not to do so. Lensen (who runs a blog critical of Google) feels it's inappropriate for Google to even *suggest* proper etiquette for *the stuff it's spent money to create*. I'm not sure why that should be so. Lensen's site, *Google Blogscoped*, states flatly, "Google wants to impose some restrictions for those books," and uses the term "impose" further down in the post. Requests are not demands or impositions. I can (and do!) request that every *Cites & Insights* reader buy at least one *C&I* book—but that's neither a demand nor at all likely to happen.

I found it interesting that, when confronted with Google's absolutely clear statement that these were suggestions, not requirements, Lensen chose to repeat that information only within the comment stream—not by modifying the post itself, as Danny Sullivan did. Anyone who reads Lensen's post and doesn't click through to comments will assume Google's up to something nefarious.

Brief notes

Terry Ballard posted "The Google E-book project: The Revolution starts now" on January 16, 2007 at *librarian on the edge* (librariansonedge.blogspot.com). He recounts an incident involving the "Library of American Civilization," 4,400 microfilm cards containing 19th century books. In 1990, he had a student check the titles against Penn's online book directory—finding a few hundred titles and adding links to the OPAC records for them. After that, Ballard had a similar project each summer. "On average, we added about 50 new titles each year, including those checked out in the summer of 2006." This time, Ballard had a student worker check the list of *unlinked* titles against GBS. "We were stunned at the number of hits that came up." At the time Ballard posted this item, another student had already added 650 new links—and they weren't through yet. Before GBS, around 10% of the LAC books were linked; that's jumped to 25%, and Ballard projects that 90% might be accessible by the end of 2007. "I can't wait to visit the Google booth at Midwinter and thank them in person."

A few weeks later, Tom Peters posted a curious entry on the *ALA TechSource blog*: "Wooden dominoes." Peters notes Princeton's addition to GLP (the 12th research university), comments on "domino theories," and seems spooked by the whole thing. In some ways, his questions—some reasonable, some less so—are less interesting than the comments, where we once

again get the absurd "privatizing" language and an astonishing suggestion that being able to search the contents of a few million books, for free, does *not* constitute a social benefit.

Jeffrey Toobin published "Google's moon shot" in *The New Yorker* for February 5, 2007 (www.newyorker.com). It's a fairly long article (ten pages printed out) on GBS and the lawsuits. I **strongly recommend** it but see no need to comment on it—it's *very well* (and I think fairly) done, as you'd expect.

It is perhaps unfortunate, if not all that surprising, that Microsoft's Thomas Rubin attacked Google in a speech at the American Association of Publishers on March 6, 2007. You can find the speech itself in Microsoft's Press Pass section. Rubin calls Google's fair use theory novel and says GLP "systematically violates copyright and deprives authors and publishers of an important avenue for monetizing their works."

"Scan this book!" in the August 15, 2007 *Library Journal* includes some of Brewster Kahle's most intemperate comments. He accuses Google Library Project of "perpetual restrictions on the public domain" and suggests Google "wants to be the only place someone can get information." This is the big lie: Kahle repeats "restrictions" later in the brief interview, even though no such restrictions are evident—and you'd think the "Find it in a library" links alone were enough to put the lie to Kahle's claim that Google "wants to be the only place someone can get information." I continue to admire the *idea* of the Open Content Alliance, but Kahle's public statements serve to damage the situation, not to help. I would hope librarians had more respect for facts and the English language.

There was a silly-season stunt at BookExpo America. Someone from Macmillan took a couple of notebook computers from Google's booth—and on returning them said, "Hope you enjoyed a taste of your own medicine" and noted "there wasn't a sign by the computers informing him not to steal them." Lawrence Lessig commented on this in a June 8, 2007 *Lessig blog* post, noting it betrayed "an astonishing level of ignorance." He offers five fundamental ways in which stealing a Google computer is different than Google's Book Search—e.g., indexing out-of-print books doesn't prevent someone from using the original. The comments—including several from copyright hardliners and a number from at least one anti-copyright extremist on "the other side"—are interesting but don't get past the fact that it was a childish stunt.

Robert B. Townsend writes a brief piece in *AHA Today* (from the American Historical Association) on April 30, 2007, "Google Books: What's not to like?"

(blog.historians.org). He cites quality-control problems with GBS and seems to proceed to trash the entire project, with this astonishing sentence in the final paragraph: “Shouldn’t we ponder the costs to history if the real libraries take error-filled digital versions of particular books and bury the originals in a dark archive (or the dumpster)?” Why, yes, we should—but incompetent actions by *libraries* constitute a whole different discussion, and for any library to regard GBS as a substitute for its collection would be grossly incompetent. Has any GLP partner suggested discarding the scanned books? Not that I’ve heard.

Google Book Search vs. Live Search Books

I’ll end this with Steve Leary’s June 17, 2007 post at *The reflective librarian* (blog.stephenleary.com). Leary says GBS claimed more than a million items as of that posting, while LSB had over 800,000. He tried using the same search terms in GBS each day for several days—and discovered that the numbers kept changing, going *down* on successive days. To me, this illustrates a general problem with Google: The raw number for a search is nearly meaningless. It does seem to be even more meaningless for GBS than for Google’s overall index. Unless GBS really is returning a page-hit count, I have no idea why. He also runs into LSB’s 250-book viewing limit.

Leary concludes, “I’m not satisfied at all with either book search product. Both refuse to give me what they promise! If I can’t see 750 books, don’t promise that many!” (In later posts, he notes that he asked people at the Microsoft and Google booths at ALA Annual about the glitches, and both said, in essence, “we’re working on it.”)

Let’s see what Leary’s phrases yield in December 2007, both in raw numbers and viewable items:

- “**Next attack**” (June: 904 [484 viewable] in GBS, 732 [250 viewable] in LSB): **GBS**: 1,038 [424 viewable] of which 705 [344 viewable] are full-view. **LSB**: 1,220 [250 viewable], of which 947 [250 viewable] are 100% viewable.
- “**Homeland security**” (June: 5,123 [155 viewable] in GBS, 749 [250 viewable] in LSB): **GBS**: 3,250 [454 viewable], of which 709 [164 viewable] are full-view. **LSB**: 914 [250 viewable], of which two are 100% viewable.
- “**Sapajous**” (June: 645 [416 viewable] in GBS, 43 [all viewable] in LSB): **GBS**: 688 [395 viewable], of which 650 [366 viewable] are full-view. **LSB**: 65 [all viewable], of which 64 [all viewable] are 100% viewable.

Conclusions? Live Search Books does a better job of representing the actual result size—but has a much

more draconian limit on what it will show you. (None of these results reached Google’s universal 1,000-result viewing limit.) As to comparative sizes and depth, other than the obvious (Google has a lot more in-print material than Live Search does), this sample size does not make any comments plausible.

Perspective

A Time of Limits?

Don’t take offense—but are you disposing of income that isn’t really disposable?

If you are, you’re certainly not alone. We seem to be in a time when “disposable income” appears essentially limitless, as long as it’s only spent a little at a time—and without much worry about the long-term consequences of short-term overspending. Part of that infinite disposability is the idea that small sums—including daily indulgences and subscriptions of one sort or another—aren’t *really* spending, since they’re so small...even as one sum piles up on top of others.

I don’t believe this is a generational thing. If it is, it’s one shared by most every generation that’s still in or nearing the workforce. It’s certainly behavior urged on us by advertisers and far too many writers who should know better. I strongly suspect there are links between the mortgage crisis and the tendency to view *all* income as disposable (including income you think you *should* have), and I hope those links may lead to rethinking of that view. Of course, maybe you’re just emulating your government, but let’s be more charitable than that.

I’m nearly certain widespread innumeracy and economic illiteracy (or aliteracy) plays a big part. I’m fairly certain perceptions of ubiquity play a role, aided by advertising and the apparent socioeconomic status of most journalists—that is, the sense that “everybody” uses X, pays for Y, regards Z as essential to life, where X, Y and Z are all uses of disposable income.

These are musings, informed by some things I’ve seen in print and in person. If I also mention the truth about some “everybody” notions along the way—well, you should be used to that by now. I may be entirely off base here. Maybe we can all keep going the way we have been with no long-term consequences—or maybe too many of us *will* keep spending as we have been, regardless of long-term consequences. But if I’m right—if some of us start to take clearer looks at spending, income, “needs” and wants, and try to bring things into balance—the consequences favor public libraries. Or at least they favor public libraries

that maintain a stance of serving the *whole* public and particularly those without lavish disposable incomes (a majority of the public by any measure I can think of), rather than slanting service toward the overadvantaged minority while proclaiming “basic service will always be free,” as basic service gets get more and more basic over time.

Consider two figures as we proceed: \$25,795 and \$48,201. Those are the 2006 median points for personal income *among wage-earners* and for household income in the United States. Remember: “median” means midpoint within a universe—so roughly half of the wage-earners in the U.S. earned *less* than \$25,795 and roughly half of the households had less than \$48,201 total income.

Does \$48,000 household income constitute middle class? And if it does, what does that say about infinitely disposable income?

Everybody Uses Netflix

Let’s start with Netflix, the way “everybody” gets DVDs. Netflix has about seven million subscribers. Blockbuster has roughly three million. (Incidentally, we *love* Netflix and have used it for years.)

“Everybody gets DVDs by home delivery.” Maybe, if there are only ten million households in the U.S. But the Census Bureau’s figure for 2006 is 116 million. And more than 80% of those households have DVD players (as of the end of 2006), and I’m going to assume that people with DVD players use them. So nearly 90% of those *who have DVD players* don’t consider even the \$9/month minimum Netflix plan to be a good value. For these people, the vast majority of Americans, convenience isn’t *that* important. At least 90 million American households have DVD players and either buy all their DVDs or have someone *go somewhere* to get DVDs—for free at public libraries or for a price somewhere else.

Are those 90 million households unimportant? They probably are to Cartier, Rolex and Lexus; they may be to Acura and Crystal Cruises. They’re not—or at least they shouldn’t be—to America’s public libraries, since they almost certainly represent at least three-quarters of your users.

What? \$17 a month (the most common Netflix subscription)? That’s *nothing*—a drop in the bucket. You’re right: For many people \$204 a year is a drop in the bucket. Why do we stick with a plan that’s only \$3 a month cheaper and means we only have two DVDs on hand instead of three? Can’t we afford \$36 a year? Of course we can. But we also watch one movie

a week and we’ve *never* had two losers in a row or two damaged movies at once—so why *should* we pay the extra \$36 a year? That’s one dinner with drinks at our good neighborhood Chinese restaurant. (It appears that the average Netflix subscriber views about 68 movies a year, based on Netflix’ own claims for shipping, so at 48 to 52 we’re just a little below average.)

Why this example first? Because some writers continue to argue that libraries need to offer Netflix-like priced delivery services in order to compete, and part of that argument is the assumed near-ubiquity of Netflix usage—a ubiquity that simply *does not exist*.

Cell Phones: Another Drop

Everybody has cell phones, right? For most of you reading this, that’s a given—and having a cell phone means having a monthly plan with hundreds if not thousands of minutes, so you can use that cell phone a lot. That usually means at least \$45 a month (including taxes and fees), or \$65 and up if data is part of the package. But that’s a drop in the bucket—\$65 a month is just \$780 a year. Now, \$780 is 1.6% of the median household income—but hey, you *have* to have a full-time subscription-plan cell phone to survive today. Of course, everybody in the household needs a cell phone, so we’re really talking \$90 to \$130 a month (or more) or \$1,080 to \$1,560 a year. That’s more than three percent of median household income—a fairly large drop by some standards. That’s *gross* household income; I’d be surprised if that translates to more than \$40,000 net, but let’s ignore taxes, Social Security and Medicare for the moment.

OK, so cell phones are reasonably ubiquitous: Market penetration in the U.S. is somewhere between 70% and 84% as of the end of 2006. Most people do have cell phones. Do they all have monthly plans? Not so much—the numbers get fuzzier, and there’s an awfully healthy market in “prepaid” cell phones.

As I’ve talked to people I consider technologically literate with decent incomes, I’m running into a fair number who have cell phones but don’t regard them as full-time communications devices. Some of us recognize that we don’t need monthly plans for those phones, particularly if we have credit cards.

We own one usable cell phone, used only for travel and special cases. We may buy another one, for cases where we both need emergency/outbound service. And until recently, we were paying \$45 a month for phone service we weren’t using.

Now we’re paying \$60 a year. If we get a second phone (on a different network), we’ll be paying \$120

a year, not the \$1,080 a year we'd be paying for two low-range monthly plans. How so? The active phone is a Virgin Mobile unit, using their "automatic top-up with credit card" plan, which only requires topping up once every 90 days at \$15 a topup to maintain the phone number. For someone who uses their cell phone a lot it's an *awful* plan: you pay 18 cents a minute (and a nickel per text message). For us, it's perfect—and if we add another phone, probably an AT&T Go phone on a similar plan, we'll be spending \$480 a year less than we were and \$960 a year less than we'd pay for a typical subscription plan.

\$960 a year almost begins to sound like real money. If I was 25 years old, I might be aware that, left in a retirement account averaging 6% a year, \$960 would be worth just over \$10,000 when I was 66 years old and ready to retire—or \$20,000 if my employer matches my contribution. That's for *one year's* savings. Do that each year and it starts to add up.

Can we afford \$1,080 a year for two full-fledged subscription cell phones? Sure—but why pay for something we demonstrably don't use? Am I suggesting that any of you cut down on your cell phone usage or subscription plans? **Absolutely not**—if you're using those plans in a way that makes your life better and you can afford the plans. I'd consider some kind of phone to be a necessity; that's why there are really cheap lifeline landline rates. (Getting rid of your landline to save \$15 or \$20 a month? More power to you—as long as there are no power outages broad enough to affect your service.)

Cable TV

Here's a case where there's much less ubiquity than usually assumed—and where I've seen more than one recent case of hard-sell journalism, writers who apparently think it's *wrong* not to have not only cable, but premium cable.

Everybody has cable? Not even close. The FCC chair recently tried to assert 70% penetration—primarily because that level of penetration would give the FCC additional power over cable operators. But according to industry figures, only 58% of American "TV households" have cable. (There are households with no TV—about four million of them in the U.S., apparently.) On the other hand, most of those households also have premium cable—but that's still a broad-minority pursuit, reaching 44% of the population.

How can you *live* without premium cable? Apparently most people think that's a reasonable ques-

tion. The average household with cable TV spends \$95 a month (\$1,144 a year) on cable service (not including broadband).

Great. No argument here. If you use premium cable as a primary source of entertainment, \$95 a month is a drop in the bucket, even if it is something over two percent of the median household income.

If you're not really using it...well, those drops start to add up. I make no claims to be a high-culture idealist. We watch *network* TV—indeed, other than our weekly movie (and old TV series on DVD on evenings when there's nothing new we want to watch), that's pretty much all we watch. We had the standard expanded-basic cable service and watched the price creep up year by year, always more than inflation, to about \$55 a month in early 2007—still less than two-thirds of the average cable subscriber's bill.

Then we looked at usage. My wife wasn't watching anything outside "basic basic" channels and I was watching one two-hour show that airs for 16 weeks a year—and since I taped it, it was really a 90-minute show. The difference between basic cable and expanded basic? \$39 a month (taxes included). Hmm. We were paying **\$19.50 an hour** for that program. That's almost enough to pay for live professional theater around here, and four times what I'm willing to pay for a TV show.

So—gasp—*we downgraded*. Saving \$468 a year. Which is a drop in the bucket—although, when added to the \$480 a year we're saving on cell phone service we don't use, it comes close to being one of those interesting four-digit drops.

I've been indirectly scolded by two sources over this rash decision. In one case, a local columnist was writing about her recently deceased mother's atrophied "spending muscle"—how she failed, according to the columnist's view, to spend enough on herself. The columnist feels the same way about her mother-in-law, bemoaning the fact that while the mother-in-law has TV and cable, she doesn't have "the movie channels." Did the columnist *ask* the mother-in-law whether she felt the need for more movies on TV? No. It seems to be a given that you're not really living if you don't have the whole package—that you're *failing to take care of your basic needs* because you don't spend \$90 a month on cable.

The other case was more insidious. A subscriber to one of the "home entertainment" magazines wrote in asking whether they actually *needed* digital cable with high-def channels when they get a high-def widescreen TV. Can't you get high-def over the air? (The subscriber was objecting to the fact that an article on

HDTV sources *omitted* over-the-air HDTV, saying “But then I guess it’s un-American not to pay for your TV reception.”)

The magazine’s writer gave a correct answer: Yes, most people *can* get network TV in HDTV form over the air, generally with an indoor antenna. The expert didn’t mention that over-the-air HDTV usually has the best picture quality of any high-def source except Blu-ray or HD DVD discs, because broadcast HDTV uses less signal compression than cable or satellite.

But the expert seemed offended by the question. “But if you’ve bought an HDTV, you want more HD content, not less, and that means paying for cable, satellite, or telco-delivered TV.” Isn’t that an interesting assumption? There’s *something wrong with you* if you don’t want more, more, more. Buy a high-quality TV and only use it for movies on disc and network entertainment? That appears to be unAmerican.

When we finally do buy a wide-screen HDTV (don’t cry for us: our 10-year-old TV is a fairly large-screen Sony XBR), we’re going to try out a good indoor antenna. If we get over-the-air reception of the high-def versions of local stations, I’ll have no qualms about booting Comcast out of our household. Sure, it’s only \$16 a month—but if it’s \$192 a year for something we’re not using, *what’s the point?* Over the course of five years, \$192 a year pays for the difference between a good HDTV and a *really* good one—or for, say, the great old TV shows on DVD.

More Drops

Warning: I’m going to mention one of those *daily* necessities and I don’t want you to get upset. You could skip this paragraph. “We all” get our \$4 Starbucks fixes every day, right? Do that once a day, six days a week, and it adds up to \$1,200 or so. Twice a day every day? Still a mere \$2,840. Plus a few hundred extra calories if you’re getting one of the coffee-flavored liquid candies Starbucks loves so much, but that’s a different issue. We’re lucky: We both used to frequent the original Berkeley Peet’s before there *was* a Starbucks, and we’ve both grown to dislike the Peet’s and Starbucks approach. I fix fresh-ground 100% Kauai coffee each morning. I’d guess it costs around fifty cents to \$1 a cup, maybe less. Not a monetary decision, to be sure: We can afford the \$2,400 a year for the household. But that’s a fairly big drop.

Then there’s bottled water—not only a cost issue, a material and fuel issue. Bottled water involves a lot of material for bottles and a lot of fuel to ship processed tap water (or spring water) from “some-

where else.” Still, since Mountain View went to chloramine instead of chlorine, the otherwise-excellent tap water tastes funny—and, unlike chlorine, chloramine doesn’t dissipate after a day in the refrigerator. We’d been buying a fair amount of bottled water (at least ours came from Hayward, a 20-mile shipping distance). We were typically paying \$4 for 24 half-liters, which is about \$1.66 a gallon for water but a whole lot better than paying more for water than for gas.

Most larger towns and cities have filtered-water stores, some using reverse osmosis (apparently the only way to get rid of chloramine). After you’ve purchased a jug, the stores around here charge two bits a gallon: \$0.25. Since the water’s purified on site and the store’s a little more than a mile away, we’re cutting fuel consumption (we fill two three-gallon jugs at a time) *and* materials—it’s easy to refill water bottles from the jugs. In this case, the environment’s more of a concern than the price: After all, at our typical drinking-water usage levels, bottled water was only costing us about \$400 a year, maybe less—which really is a drop in the bucket. Of course, \$52 to \$78 a year is even better.

Disposing of Income

I could go on, particularly for those enamored of the “celestial jukebox” concept. \$10 a month for satellite radio? Nothing. \$13 a month for a music service? Nothing. \$50 for *really* high-speed internet? Don’t be silly: That’s nothing. (We spend \$20, but that’s because we’re too far from the switching office for higher-speed DSL and don’t want to deal with Comcast.) Other daily or frequent indulgences? Well, we all need our indulgences (although I find that infrequent indulgences are somehow more satisfying than daily indulgences). I don’t argue against them. Life is tough; indulgences can help. Shopping out of boredom? Ah, that’s another story...

When you add all those nothings together, it can start to seem like real money. For a two-person household where both buy coffee drinks daily, with two cell phones, premium cable, Netflix and satellite radio (but no by-the-bottle bottled water), I come up with nearly \$6,000 a year—12% of a median household’s gross income. Which is fine—if the couple actually has \$6,000 a year in *disposable* income.

There are “necessities” and then there are **necessities**. Maybe you need Netflix, premium cable, coffee drinks, bottled water, one cell phone per person and all that—but it may be a different level of need than for housing, water, electricity, gas, food, clothing,

transportation and health insurance or by-the-appointment medical care. None of those needs is glamorous, but they make up a sizable chunk of income that's really *not* disposable—a big chunk for most households earning much below, say, \$75,000 a year (and quite a few above that level). And 70% of U.S. households are below that level.

What's disposable income? To advertisers and merchants, it's anything they can get you to spend. I'll offer a few counter suggestions:

- If you're not contributing the maximum for matching funds to retirement, you may be overspending your disposable income.
- If you have outstanding credit card debt, you may be overspending your disposable income; if you're using a measurable percentage of your income for credit card interest, you're *probably* overspending disposable income.
- If you're paying more than 30% of your gross income for housing—mortgage and utilities—there's a good chance your disposable income is at risk. If you're paying 40% or more and you have an adjustable-rate mortgage or a trick mortgage...well, you probably already know you're in trouble.

When Did We All Become Upper-Middle Class?

What's been happening the past few years seems to be predicated on the idea that everyone has loads of disposable income. In essence, we're all upper-middle class or better, since in the past you could assume that middle-class folks made hard decisions about their money. Somehow, though, we're now in Lake Woebe-gone country—we're all above average, at least in disposable income. And we *need* all those things our disposable income buys.

One article, pushing the idea that public libraries should be *selling* premium services such as Netflix-style home delivery, says most public library users have incomes between \$15,000 and \$35,000 a year, and argues that libraries should target higher-income users. (There's a significant flaw in *any* non-Census-Bureau survey these days that purports to include income as a characteristic: Many of us refuse to answer such questions, either choosing "Prefer not to state" or abandoning the survey if necessary. My *guess* is that higher-income people are more likely to consider their incomes private.) Implicit in that argument—at least to my eyes—is the idea that most people are well above that income range. Are they?

Not according to the census bureau. Yes, a lot of Americans make more than \$35,000 a year—but "a lot" is roughly one-third of those over 15 years old, or about one-quarter of all U.S. residents. (If you're wondering, roughly 11% of wage earners earned more than \$70,000 a year in 2006.)

What would you consider upper-middle class or affluent? I'd suggest a household income of \$150,000 a year or more, at least in high-cost states. By that measure, only 7.5% of households qualified last year—just under 8.8 million households out of 116 million total. Let's use "comfortable" rather than affluent. I'll suggest two rough levels for "comfortable"—\$100,000 for high-cost areas, \$75,000 for the nation as a whole. By those standards, a lot of us *are* comfortable, but still a minority: about 19% of households in the former category, about 30% in the latter.

Should public libraries be targeting services for that 19% or 30%? Maybe, at least to some extent. For the 7.5%? Certainly not at the expense of other services. It's hard for any two-tier service program with charges for the upper tier not to distort the library's functioning to some extent—on behalf of those *least in need* of commonly-supported goods. Your answer to the 7.5% question may have a lot to do with whether you regard public libraries as a public good or as another cultural outlet for the moneyed.

Living the American Nightmare

Numeracy plays a role here. So, to be sure, do advertising and predatory lenders. Too many people purchased houses they could not afford unless everything went right. (Yes, some of this may have been fraud—but how can someone bringing home \$2,500 a month *not* see that signing mortgage documents to pay \$2,600 a month—omitting utilities, food, clothing, transportation, etc.—is asking for trouble even *before* the rate adjusts? That's the kind of story we're seeing here in some cases, and it's hard to buy "Oh, they were just misled" as a total story.) Too many people with modest incomes bought into the idea that they deserve everything, and they deserve it *now*.

For most of us, I think, this isn't about austerity. It's about choices. Sure, you can have that BMW—but it may mean you won't have as much in retirement savings. You can *certainly* have those coffee drinks, that premium cable, Netflix and whatever—but maybe not all of them, not all at once, adding new indulgences and monthly services as you go. Maybe you keep your car for five years or seven years instead of three. (Or ten years or twelve years...fact is, I still

think of my six-year-old car as a new car, and probably will for years to come. Admittedly, we live in California, where rust on the undercarriage isn't much of a problem—and, by the way, there are a lot of ten, twelve, 15-year-old cars on the roads here.)

Want that great cruise? I'm all for it. When we went on one trip and saw people who were then our elders having trouble with some of the excursions, we decided to go places while we were still sure we could fully enjoy them...and, later, discovered the virtues of high-end cruising. Those were expensive choices, and for us they were the *right* choices. We chose those cruises. We didn't choose some other expensive purchases. We didn't choose to frequent high-end restaurants all the time. We didn't choose expensive clothes. But we certainly chose to spend serious money—making choices and using true disposable income.

But choices don't seem to be the American way. What does seem to be the American way: Ignoring tomorrow. That, I believe, is changing—or at least I hope it is. Maybe it's because I'm nearing retirement age, but I'm becoming acutely aware of the fiscal realities of retirement. Just a brief digression:

Traditional advice has been that you need 80% or so of your current net income once you retire. Increasingly, though, that's nonsense. What costs are really going to go *down* after you stop working? In many households, you'll spend *more* on utilities, since you'll be home much more of the time. Health care? Maybe you'll spend less (with Medicare); maybe you'll spend more. Otherwise—well, if you've been spending a fortune on professional clothes, sure, there's a savings, but most of us haven't been. Food? Could be cheaper in retirement (if you've been going out for lunch a lot). Could be more expensive. Meanwhile, you'll probably want to go on more vacations. You'll probably take up more hobbies—some expensive.

I'm seeing contemporary advice that assumes you'll spend as much after retirement as you do before. That still doesn't require as much gross income, to be sure, but it's a sobering thought. Add to that the 4% guideline: Your retirement savings will last you long enough, accounting for inflation, if you start out using *no more than 4%* of your gross savings.

If those numbers are sobering, they should be. When I read about a 55-year-old couple who've saved half a million dollars and decide to go for early retirement, I wonder what they plan to do if either one reaches 75 or 85. Half a million dollars sounds like a lot of money. \$20,000 a year—not so much.

That isn't what we want to hear. I was reminded of the economic fantasy we're encouraged to lead by a

December 5, 2007 post at *Stephen's lighthouse*, “8 consumer trends” (http://stephenslighthouse.sirsidynix.com/archives/2007/12/8_consumer_tren.html). Stephen Abram points to a *Trendwatching* report on trends for people to “watch and capitalize on in the new year.” Let's look at some of them:

Status spheres

We get an inkling right up front of where *Trendwatching's* coming from, given this quotation from their report from the start of this year:

[I]n the end, when dealing with (and selling to) people, everything always comes back to status. In a traditional consumer society, he or she who consumes the most, the best, the coolest, the most expensive, the scarcest or the most popular goods, will typically also gain the most status.

We're all out for Status, which means lots of spending. But even as consumption-oriented an outlet as *Trendwatching* recognizes that sheer consumption doesn't cut it any more, so they talk about “*Status Spheres*: a variety of lifestyles, activities and persuasions, which can be mixed and matched by consumers looking for recognition from various crowds and scenes.”

They count on the “traditional sphere,” with “**hundreds of millions** of consumers who *do* want to consume more, who *do* covet all things bling, who *do* crave in-your-face brands,” but there are more “spheres,” such as the “transient sphere” for bored folks “driven by entertainment, by discovery, by fighting boredom, who increasingly live a transient lifestyle, freeing themselves from the hassles of permanent ownership and possessions.” [Emphasis added: really? *Hundreds* of millions?]

Or the “online sphere,” where “social status 2.0 is all about who you connect to and who wants to connect to you, tribal-style.” This group's relentless commercializing of everything even manages to cheapen the “eco sphere” (people who care about the environment), the “giving sphere” (philanthropy), and the “participative sphere.” Ah, but look:

Premiumization

Here's the short version: “with more wealth burning holes in (saturated and experienced) consumers' pockets than ever before, quick status fixes derived from premium products and premium experiences will continue in full force next year.”

We all have too much money and are out to get “quick status fixes.” How do we do that? How about \$20 bottles of water—or, for that matter, “Bling H2O” at \$480 a bottle. Limited-edition beer at the price of Champagne. \$28 gourmet marshmallows.

How about “fashionable toilet paper” at \$4 a roll? Porsche baby strollers? \$5,000 leatherbound laptop computers? The report fails to distinguish between absurdity and actual differentiation, but it’s clear that the keys are *expensive and exclusive*—regardless of worth.

Snack culture

You’ve probably heard about this before: Breaking stuff down into bite-sized pieces—so you can consume an ever-wider variety without noticing when you’ve gone too far. Hundred-calorie packs of snacks cost a lot more per ounce than larger packs—but they’re worth it, apparently. Small-plate restaurants with big-plate prices are all the rage. How about semi-disposable clothes, intended to be worn a few times and then discarded? That doesn’t do a whole lot for the environment, but the prices seem low. *Wired* has celebrated the rise of “snack culture” and the apparent concomitant ubiquity of short attention spans.

Each tidbit costs less than the big chunks you used to buy. So what if you’re spending twice as much overall? You’ll never notice along the way...

I won’t go through the others. The sheer cynicism of the “Eco-iconic” section and assertion that we all have an “insatiable demand to be online 24/7” are annoying enough; a couple of the sections are, well, just plain strange.

But that’s also what’s being peddled—by *Wired*, by many consumer magazines, by loads of specialty cable channels (and to some extent by all ad-supported TV). Consume more, consume “better” (higher status, more expensive) and don’t worry about the consequences.

A Time of Limits?

Are there limits? If so, will more of us come to recognize them? To bring in another long-time theme, will we seek lives in balance?

I hope so. I’d like to think so. I’m not arguing for budgeting (unless your spending really is out of control). For many of us, that’s a needless annoyance. I’m *not* telling you to change your ways—unless your ways are causing you to lose sleep or worry about your ability to sustain your lifestyle.

Lifestyles are overrated. There’s a difference between maintaining a lifestyle and living a good life. One is a matter of recognition, status, consumption; the other is a matter of balance and inner peace. It’s tough to maintain a given lifestyle if your income slumps a little or you have unexpected expenses: Those daily “needs” hurt when they’re gone. It’s easier to keep living a fulfilling life when your circumstances

change slightly. Living within your limits can be *good* living, even if it doesn’t match an assumed lifestyle. There’s a funny thing about living within limits and paying less attention to status: You may find that you have *more* disposable income for things that would improve your life, even if only for an hour or two.

People who live within limits are more likely to make good use of shared assets, I suspect. They’re more likely to appreciate parks, to take walks...and to use their public libraries. I’m hoping more people will recognize the need for limits without having that need forced upon them through foreclosure or bankruptcy or an inability to retire...ever.

Masthead

Cites & Insights: Crawford at Large, Volume 8, Number 1, Whole Issue 98, ISSN 1534-0937, a journal of libraries, policy, technology and media, is written and produced by Walt Crawford, Director and Managing Editor of the PALINET Leadership Network.



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

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

Comments should be sent to waltcrawford@gmail.com. *Cites & Insights: Crawford at Large* is copyright © 2008 by Walt Crawford: Some rights reserved.

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

URL: citesandinsights.info/civ8i1.pdf