

Cites & Insights

Crawford at Large/Online Edition

Libraries • Policy • Technology • Media

Volume 12, Number 7: August 2012

ISSN 1534-0937

Intersections

It Was Never a Universal Library:

Three Years of the Google Book Settlement

Remember the Google Books settlement? It was going to settle a four-year-old pair of lawsuits (four years old then, eight years old now) against Google (by the Association of American Publishers, AAP, and the Authors Guild, AG) asserting that Google was infringing on copyright through its two-line snippets from in-copyright books scanned in the Google Library Project—and by the scanning itself. Later, a third group representing media photographers also sued Google for the same actions.

Sections

General Notes: Before the Outcome.....	2
General Notes: After the Outcome.....	19
Orphan Works	37
Monopoly and Antitrust	46
Privacy and Confidentiality	52
The Public Domain, Open Access, Copyright and Fair Use.....	58
Libraries and Metadata.....	63
Authors and Publishers.....	97
Class and Standing.....	101
The Future	110
In Closing	118

A proposed settlement was announced in October 2008. Lots of people had lots of things to say about it—not unreasonably, since it had major implications. The [March 2009 Cites & Insights](#) is a 30-page discussion of the settlement and what was being said about it. An [essay in the July 2009 issue](#) addressed the misuse of the English language by some commentators. I assumed—as I believe most other observers did—that the settlement might be modified slightly but would probably be approved within a year or two, maybe even faster than that.

Now? The settlement (modified) is dead: The judge struck it down as being unfair. Most of those who were commenting on it (including me) really didn't deal with what turned out to be the core issue: You can't substantially transform copyright law by settling a class action lawsuit.

We are, in some ways, back to square one after the better part of a decade. There will assuredly be more developments over the next (year? five years? decade?), but given the clear death of the settlement itself, I thought this would be a good time to update the situation.

If you've managed to ignore the settlement (called GBS for convenience, as it is by at least one of the truly knowledgeable commentators) so far, I'll suggest reading my March 2009 overview and possibly a few of the items it points to. I'm not going to rehash it—as it is, this discussion is longer than the earlier one, even as it's fundamentally a story of failure.

Or is it? Maybe the failure of GBS is a success in other areas—including (potentially) areas such as fair use and sensible planning for library futures.

This is a long set of notes and comments (cites & insights). It strikes me that the topic and complexity deserve that length—but note that I'm offering much *brief*er excerpts and comments on most items than I normally would in this sort of roundup.

After two sets of general notes and overviews (one before the settlement was rejected, one after) I'm breaking the discussion down by topics rather than chronologically.

General Notes: Before the Outcome

It may be amusing to start with the single item I retagged “gbs-paranoia” when I was retagging nearly 300 “gbs” items in Diigo. It's by Steven Levy, [posted at wired.com](#) on March 31, 2009 with the title “Who's Messing With the Google Book Settlement? Hint: They're in Redmond, Washington.” It's as fair-minded, balanced and objective as most of Steven Levy's writing, especially where certain computing companies are concerned.

Here's the nub of the “story”: New York Law School's Institute for Information Law and Policy filed an amicus curiae brief (or, when Levy wrote this, *planned* to file a brief) during the pre-hearing period set for such briefs—as did many other parties.

Explaining what the New York Law School brings to the party, [Daniel] Kornstein cited its mission “to understand the interplay of law and technology and influence their development to serve democratic values in the digital age ... to extend human knowledge and harness new informational tools to the goals of social justice.”

The Institute, he writes, “is in a position to make a significant

contribution to the resolution of the legal issues in dispute by virtue of its recognized scholarly expertise in intellectual property and Internet law.”

Which seems reasonable—but apparently it’s *not* OK (at least in Levy’s mind) for Microsoft to underwrite that contribution.

The chief investigator of the New York Law School project is James Grimmelmänn. In an earlier career phase, associate law professor Grimmelmänn worked as a programmer for Microsoft. At a conference in February, Grimmelmänn was discussing his views of the book settlement with a policy specialist of his former employer, and the Microsoft exec reminded Grimmelmänn that the company has had a continuing interest in funding academic efforts.

Microsoft provided \$50,000. According to a Microsoft counsel, Microsoft funds “dozens of law projects.” Microsoft had no say in the content of the brief. Frankly, I know of *nobody* other than Steven Levy who regards Grimmelmänn’s GBS work as being biased or less than first-rate: He’s generally acknowledged as the go-to commentator. But here’s Levy’s final paragraph in this nasty little hit piece:

Turns out that cleverest hacker here is Microsoft, making an academic grant that may help put some judicial heat on its rival.

Now, on to writers who are less into Heroes & Villains as standard operating mode.

The Audacity of the Google Book Search Settlement

A striking title for [this August 10, 2009 piece](#) by Pamela Samuelson at *Huffington Post*. Samuelson is a law professor at UC Berkeley. Here’s the opener:

Sorry, Kindle. The Google Book Search settlement will be, if approved, the most significant book industry development in the modern era. Exploiting an opportunity made possible by lawsuits brought by a small number of plaintiffs on one narrow issue, Google has negotiated a settlement agreement designed to give it a compulsory license to all books in copyright throughout the world forever. This settlement will transform the future of the book industry and of public access to the cultural heritage of mankind embodied in books. How audacious is that?

She recounts the two lawsuits briefly and notes a couple of key points, after noting Google’s claim that the snippets constituted fair use and that Authors Guild did not fairly represent the class of affected authors:

Many copyright professionals thought Google had good defenses on both issues. Google's attack on Mitgang and the Authors Guild as class representatives would likely have succeeded because most authors of books in the Michigan library are academic researchers likely to think, as I do, that scanning books to make indexes and snippets is fair use. There are approximately 100 times more academic researcher-authors than there are members of the Authors Guild.

I like (and find hard to fault) Samuelson's somewhat cynical comments on why Google, the AAP and the Authors Guild were all willing to settle:

So why did Google decide to settle instead of to fight? Inspired perhaps by Rahm Emanuel, who has observed "you never want a serious crisis go to waste," Google recognized that AAP and the Guild would be willing to settle their lawsuits by vastly expanding the plaintiff class to all persons with a U.S. copyright interest in one or more books. The settlement could then give Google a license to commercialize all books owned by the class.

Why would AAP and the Guild be willing to do this? It is largely because the agreement designates the Authors Guild as the representative of the author subclass and the Association of American Publishers (AAP) as the representative of the publisher subclass. This designation ensures that they will have vastly expanded responsibilities and powers to control the market for digital books for which they have been hankering for many years.

After further discussion, Samuelson focuses on the non-representativeness of the Authors Guild as one reason to object to the proposed settlement—noting that the terms serve the interests of AG and AAP members much better than they do "the thousands of times larger and more diverse class of authors and publishers of books from all over the world." *Thousands* of times larger? Yep: AG has about 8,000 members; she cites OCLC estimating 22 million authors of U.S. books since 1923—and AAP is essentially the Big Six, while there are tens of thousands of small publishers in the U.S. and abroad. It's a good brief comment on one good reason to question the settlement from a respected source who's on the skeptical side of the fence.

Pros and cons of the Google book deal

I suspect [this May 1, 2009 piece](#) by David Weinberger appearing in *KMWorld* offers a fairly typical attitude as to what was likely to happen—an attitude I shared at the time, based on the reading I'd done:

There's no particular reason to think it won't go through, although many people are objecting to various parts of it.

Given that assumption, Weinberger phrases the opening not as a "could be" but as an "is":

The Google Book Search settlement is huge, complex and overall a big step forward. But it's also quite scary. The world of print is about to change, mainly for the better.

What he believes to be the good points? The first paragraph of that section gives me pause, partly because I think of "indexing" differently than what Google does with books:

The publishers are likely to make submitting their books for indexing a regular part of publishing. That means that we'll be able to search them via Google, see a preview and press a button to buy a copy. Books that are out of copyright will be fully readable and downloadable for free, as is only proper.

He regards the OP/orphan works portion as the most significant "goodness." Additionally, he's enthused about being able to do text analysis over the entire corpus of Google Books.

But, for all this joy, there are big, worrisome issues, mainly because this is a settlement between Google, authors and publishers. Can you think of people whose interests are not directly represented in this agreement, hmm? Readers, perhaps? Scholars? Educators? Libraries?

He finds three objections "especially trenchant:"

- The supposed *de facto* monopoly on scanning, indexing and accessing books—and here, Weinberger makes what I regard as an extreme leap: "Google is about to become our national library."
 - The second "objection" is odd: "Second, the settlement should clearly maintain at least the old standards of Fair Use. We don't want to end up with even less ability to reuse our culture than we had before. The existing settlement is a lost opportunity to clarify and expand Fair Use." While I agree with the final sentence, there was nothing in GBS that *narrowed* fair use. It simply didn't address it; Google retreated. I can't see raising fair use as an objection to GBS itself.
 - Institutions will be charged for accessing the digital library.
- He notes other issues briefly. What I find most interesting here is the assumption that the deal *will* go forward and the (to me) odd set of objections raised.

Google's digital-book future hangs in the balance

I'm citing Stephen Shankland's [June 15, 2009 piece](#) at *CNet News* because it's a reasonably good, reasonably brief overview of (some of) the issues around GBS, quoting a number of those arguing for and against it.

Not that it's perfect. I could have done without the description of physical libraries as "musty archives" and the alternative wording: "If the company succeeds in its ambition, the world's books will emerge from dusty library stacks to be reborn on the Web, and Google already has a 7-million book start." It *must* be possible to refer to libraries without labeling them musty or dusty or some other term implying that nobody would (horrors!) actually use them, but, hey, Shankland's a tech writer for a tech site.

One sentence is either hopelessly naïve or just wrong: "Though search is Google's primary business, the company also stands to make money directly from book search." Search is *not* Google's primary business. Advertising is Google's primary business. Search is one way Google sells advertising. (A company's primary business is what it makes the most money doing.)

And then there's a quote from law school professor Randal Picker that indirectly, to my mind, says that the settlement didn't make legal sense from the beginning (in ways I didn't grok at the time):

"What I think the judge needs to think about is whether we think **the Authors' Guild would on its own grant a similar license** to competitors to Google. If answer is no, and there is good reason to think they would say no, this license will by its terms create monopoly power," Picker said. "There is a chance this is the only orphan-works license that will be created. No one else like the Internet Archive would be in a position to compete with Google with respect to the orphan works." [Emphasis added.]

I suddenly say to myself: "Who gives the Authors Guild, representing 8,000 authors out of millions, the authority to grant a license of such scope in any case?" I can't think of a satisfactory answer. For AG to claim authority to grant a compulsory licensing scheme for orphan works, at least 99% of which were written by people who are not members of AG (that's a guess, but it's an educated one), is simply absurd.

Google Books and the Judge

This piece by Anthony Grafton [appeared September 18, 2009](#) on the *New Yorker* "Page-Turner" blog. The first paragraph is a newsy item that I believe overstates the impact of its topic: That is, an agreement that Google would "allow" On Demand Books to produce paperback versions of public domain (the piece calls them "out-of-copyright") books using the Espresso Book Machine at a recommended \$8 price. "The Google-On

Demand partnership could transform retail bookselling—especially of books for university courses.” Really? Making *two million* books all published prior to 1923, all of which are freely available for the taking, printable via POD at a fairly high per-copy price, can *transform* retail bookselling? “Especially of books for university courses,” since most courses rely *so* heavily on public domain materials? Who knew?

That’s snark. Sorry. That leads into the real story, perhaps: That September 18, 2009 was the deadline for submissions to the court regarding GBS.

The settlement has a lot to offer most ordinary authors—those of us whose books sell in the high hundreds or low thousands, and then go out of print. Google will pay sixty dollars for every book for which it can find a rights holder and will share any future revenues with authors and publishers. More important, millions of books that are in copyright but out of print (and hard to find) will get another chance. People searching for information will learn from Google that these books exist and then be able to read sections of them online. The system will provide immediate links to libraries where the full texts can be found and to retailers, if any, who sell them. Any rightsholder who doesn’t want to take part can opt out.

From most writers’ standpoint it looks like a decent deal.

Note that we’ve jumped from two million public domain books to the millions of out-of-print books, with no recognition that these are entirely different groups. After that enthusiastic paragraph, Grafton notes some of the problems—e.g., complaints from the Register of Copyrights, complaints from France and Germany, Amazon’s “predictable” complaint, trustees of writers’ estates...and “even the libraries that have provided Google with its raw materials.” Grafton also talks about metadata issues, quoting Geoffrey Nunberg (see later in this roundup) and doesn’t come to any conclusions. It’s an odd little news story, conflating two very different topics in a way I find unconvincing. He assumed a decision would be reached in 2009—“Will the juggernaut keep rolling? We’ll know later this year.” It’s not only those *not* in the know, like me, who badly underestimated how long things would actually take.

Google Book settlement: Alternatives and alterations

I believe this perspective by John Mark Ockerbloom, [posted September 17, 2009](#) at *Everybody’s Libraries*, is the last of these overviews that deals with the *original* GBS. Ockerbloom was pro-settlement: he feared that a collapse “might deprive the public of meaningful access to millions of out-of-print books.” This post is about alternatives others have suggested, along with

Ockerbloom's explanations of "why they don't seem to me as likely to succeed on their own." He discusses four possibilities:

- Compulsory licenses similar to those in songwriting—and in some odd ways GBS would establish a compulsory license of sorts. He notes that the settlement could be modified such that equivalent licenses *had* to be made available to others, but also that Congress' general tendency is such that it would be unlikely to pass a compulsory license law. (Ockerbloom notes that, while the Copyright Office notes such licenses, it "tries to damp down the idea" and characterizes licenses as happening only when there's clear marketplace failure.)
- Orphan works legislation. "An orphan works limitation on copyrights would be nice, but it's not going to enable the sort of large, comprehensive historical corpus that the Google Books settlement would allow." That's true—it wouldn't create a "near-comprehensive library of millions of out-of-print 20th century books" because many of those millions are not *at all* orphans. He also notes that a 2008 orphan works bill was abandoned by Congress because groups of copyright holders objected.
- Private negotiations between Google (or "other digitizers") and each rightsholder. Possible for the Big Six; impractical in general.
- Copyright law reform. Here I'll quote Ockerbloom's comment, which covers it fairly well:

As James Boyle points out, it would solve a lot of the problems that keep old books in obscurity if books didn't get exceedingly long copyrights purely by default. It would also help if fair use and public domain determination weren't as risky as they are now. I'd love to see all that come to pass, but no one I know that's knowledgeable on copyright issues is holding their breath waiting for it to happen any time soon.

Ockerbloom was among those who regarded GBS as imperfect but "still the most promising starting point for making comprehensive, widely usable, historic digital book collections possible." When you read this commentary, also read the handful of comments—including jrochkind's, since that commenter has the same understanding that I do: effectively, compulsory licensing allows for "cover versions" of a song you wrote (by paying you or a licensing agency) but doesn't mean I can start copying and selling *your performance*.

The Google Books Settlement: Who Is Filing And What Are They Saying?

Brandon Butler prepared this for ALA, ARL and ACRL [on September 28, 2009](#). It's a nine-page PDF summarizing key information about "the

hundreds of filings that have been submitted” regarding GBS. Most of the summary is a few pages of tables.

After a table showing how many filings there were—more than 400 if I count correctly, but nearly 300 of those are foreign agencies objecting to inclusion in the classes—and some brief tables summarizing key objections and support elements, there are tables showing key supporters (filing number, party, reason for support), filers “with reservations” (a category that includes ALA/ARL/ACRL, AAUP and others) and key opponents (the longest list, and one that includes The United States of America). Well worth checking if you want to explore this in depth.

The Long and Winding Road to the Google Books Settlement

[Jonathan Band’s article](#) in the *John Marshall Review of Intellectual Property Law* 227 (2009) is a key document for those wishing to understand the GBS story in depth. It provides a clear history of the lawsuit, the initial GBS and some of the objections raised. It notes that the Department of Justice, on the last day for filings, basically recommended that the settlement be turned into the status quo, which Band found lacking:

In other words, the United States encouraged the parties to take the Library Project back to where it started: an index with snippet displays of search results. The institutional subscription and consumer purchase would be available only with respect to books whose rightsholders had opted-in for such access. Observing that Google had suggested that the vast majority of known authors and publishers of out-of-print works who had received notice of the settlement would wish to be bound by it, the United States opined that “creating an opt-in mechanism would not seem to work a significant hardship for a broad category of affected works.” This is a complete non-sequitor. Google’s belief that most known rightsholders would not oppose the settlement does not mean that both the known and the unknown rightsholders are likely to opt-in to an electronic distribution system. Given the small amount of probable compensation, many rightsholders might not bother to file claims with the Registry. Moreover, because most of these books currently have no economic value, the heirs of the authors of many of these books do not even know that they are rightsholders. Accordingly, an opt-in institutional subscription database would probably be far less comprehensive, and thus far less useful to serious research, than the institutional subscription database proposed under the settlement.

At that point, the parties involved asked Judge Chin to cancel the fairness hearing and went back to negotiations, emerging on November 13, 2009

with a revised settlement which has been called GBS 2.0 (and various other formulations). Key changes (discussed in much more detail in Band's article):

- The agreement for full-text display and other services beyond snippets would leave out books *not* published in the U.S., Canada, UK or Australia unless they'd been formally registered for U.S. copyright before January 5, 2009—probably eliminating half the books.
- The Registry would have publisher and author representatives from each of the four nations.
- Instead of holding revenues for “unclaimed” works (essentially true orphans) for five years, then using them to cover Registry expenses, the held revenues would be used to search for orphan-works authors and for literacy-based charities.
- GBS2 allows for renegotiation of revenue splits for commercially available books and changes some deadlines for opting out.
- A number of changes would make GBS2 *slightly* more open to competition. There are also some other changes in detail and one possibly major change: An explicit waiver of a possible claim that GBS immunizes its parties from antitrust actions.

At that point, the new timeline was supposed to result in a February 2010 fairness hearing. Band didn't think that would be the end of the road, even if it had taken place then:

Of course, even if the court approves the ASA, the case is far from over. Class members can appeal the court's decision to the Second Circuit. Likewise, if the court rejects the ASA, the parties can appeal that decision to the Second Circuit. Moreover, foreign rightsholders excluded from the ASA could bring copyright infringement actions against Google for scanning and displaying snippets of their works. In short, the long and winding road to the Google Books settlement is far from its ultimate destination.

Once again: This is a key document, one I highly recommend for those wishing to understand the GBS through November 2009. Band writes well and (to some extent) from a library perspective. Why don't I just say “go read it; I'll wait”? Because it's a 104-page (8.5x11) PDF—and even though perhaps 1/3 of that (maybe more) is taken up with 937 footnotes, that's still a fair amount of reading. Obviously, I haven't really attempted to summarize!

Google Books Settlement 2.0: Evaluating the Pros and Cons

This piece, [posted November 16, 2009](#) on Electronic Frontier Foundation's (EFF's) *Deeplinks Blog* by Fred von Lohmann, is the first of

several EFF posts evaluating GBS2 (which we'll just call GBS or GBS2 most of the time from here on).

When it announced its Book Search project in 2004, Google set for itself an inspiring and noble goal. In the [words](#) of Google CEO Eric Schmidt, “Imagine yourself at your computer and, in less than a second, searching the full text of every book ever written.” What started as a dream of universal book search, however, has become something much broader: a class action lawsuit and proposed settlement that hopes to let Americans read, as well as search, millions of books online.

Instead of offering one very long discussion, EFF's take is broken down into several relatively brief parts—this post, for example, is only five paragraphs long, although those that follow are longer. Von Lohmann recommends Grimmelmann's *Laboratorium* as a good ongoing source. Here's the key paragraph for this brief introduction to a series of posts (some noted later in this section or elsewhere in this article):

Here's a preview of the overall contours of the debate. The chief benefit of the proposed settlement is the increased public access to books (particularly out-of-print books) that it makes possible. Against this important benefit must be balanced concerns about possible detrimental effects on privacy, competition, innovation, and fair use. Complicating the overall analysis are the requirements and limitations of class action litigation, as well as the inherent difficulty in predicting how copyright owners and readers will respond to the new Google products and services contemplated in the proposed settlement.

Google Books Settlement 2.0: Evaluating Access/Evaluating Censorship

These two continuations of the item above, by Fred von Lohmann at *Deeplinks*, appeared on [November 17, 2009](#) and [December 3, 2009](#) respectively.

The first is mostly about potential upsides of GBS2: enhanced public access and unprecedented online access (at least in the U.S.). It's a good, brief, *fair* discussion (as far as I can tell). But it's also about “The Uncertainty: Empty Promises, Empty Shelves?”

First, under the settlement copyright owners can pull their books (see Section 3.5, “Right to Remove or Exclude”) out of all the products and services envisioned by the settlement, including full-text search and limited “snippet view” access. This is essentially the “take the money and run” option—the copyright owner collects a

per-book payment from Google for books already scanned, but then the public gets no online access to these books unless and until the copyright owners negotiate new deals with Google or other online providers. This effectively gives copyright owners a unilateral right to trump fair use, essentially “unpublishing” their books online. [Some observers](#) expect that most major publishers will opt to “take the money and run” for both their in-print and out-of-print titles, leaving gaping holes on the virtual shelves of Google Books. If this takes place, then the settlement would only foster access to orphan and unclaimed works. Still good, but far short of full access to every book in the University of Michigan library.

Then there’s the fact that Google isn’t *required* to offer all the products and services it’s *allowed* to offer. That seems to be less of an issue than the third problem:

Third, the public gets only the kinds of access that Google makes available, only through interfaces that Google chooses to expose. And while this level of access is certainly preferable to no access at all, the “One Interface to Rule Them All” approach is likely to impede innovation, which ultimately means less access. It would be preferable if others had access to the underlying book scans, just as Google had access to the World Wide Web when it built its own search engine. (Google will protest that it spent the money to make the scans, and it’s unfair to allow competitors to free-ride on its scanning investment. We already posted our [answer](#) to that objection.)

That’s a good point I hadn’t really seen elsewhere (probably due to inattention).

The second piece speculates on the forms of censorship that could take place within the digital corpus—using a somewhat broad definition of “censorship,” since the books within the GBS service still exist in physical copies and certainly haven’t had publication prevented by the government. (I think von Lohmann’s usage of “censorship” for what he’s discussing comes very close to being language abuse on the order of “privatizing,” but let’s set aside that tedious argument.) He sees three categories of risk:

- **Censorship by rightsholders:** Copyright owners can make their works wholly invisible within Google Books—that is, neither viewable nor searchable. There’s much of this “Last Library” nonsense to suggest that a book dropping out of Google Books means *it no longer exists*, which is bull. von Lohmann also decries the possibility of

editing—but his solution, “a prohibition on *anyone* making editorial alterations in the text of scanned books,” has the effect of precluding cleanup efforts on the sloppy scanning.)

- **Censorship by Google:** The settlement “gives Google a troubling degree of discretion when it comes to choosing which books will be publicly accessible.” I hate to sound like an advocate of private enterprise, but typically a private company *does* have some discretion in deciding what it will sell. (Again, this is no more censorship than the first is, since *the books are still there*.)
- **Censorship by government:** “Finally, it’s worth noting that governments will doubtless exploit the leeway that the settlement gives to both rightsholders and Google to pull books off the digital shelves of Google Books.” Again, this *would not be censorship*, but it comes a little closer.

And, of course, Google could sell off the whole project. Well, yes it could; otherwise, Google ceases to be a private company.

I must admit that I find the second essay unconvincing—largely because *none* of this is censorship unless you stipulate that physical books are going to disappear as soon as Google starts up the so-called “Last Library.” I’m not willing to make that stipulation.

Nitpicking the Google Books Settlement 2.0

That’s Gavin Baker posting [on November 18, 2009](#) on his eponymous blog, focusing on points he regards as salient that he doesn’t think have received much discussion. He notes the loss of most international works and says he has seen *no* criticism of this loss of access (but it was at this point only five days after GBS2 was posted). He notes, properly, that saying “foreign language works are now excluded” is wrong on both counts—some foreign language works would be included (either because they were registered in the U.S. or because they were published in the U.S., Canada, UK or Australia) and some English-language works would not be.

On the other hand, he doesn’t buy the criticism of orphan works provisions, since he sees access to orphan works as the biggest benefit of the settlement.

The main criticism of this is that Google would be the only provider of access to these orphan works. Monopoly access is certainly undesirable (particularly given the other flaws of the settlement: the privacy weaknesses, the DRM, the single interface, the overall market position of Google, etc.). But isn’t monopoly access (with antitrust scrutiny) better than no access?

The only way the answer is “no” is if the settlement holds back progress toward non-monopoly access. For instance, a settlement clause that guaranteed Google competitors the same terms (even if they had to do the scanning themselves) would open competition. Obviously, Google is not interested in such an approach, and since the settlement is a negotiation between Google and the plaintiffs (who I would guess to be agnostic on that question), we shouldn’t expect to see those terms unless the judge or the Department of Justice forces them.

I’m not sure you can ignore the monopoly issue that easily, although—at least in 2009—I think I would have agreed with Baker. I do, unfortunately, agree with him that legislative progress on orphan works is unlikely.

He also discusses limitations in the powers of the Unclaimed Works Fiduciary, the independent agent to manage what are effectively orphan works (any works not claimed by rightsholders): To wit, it can only exercise normal rightsholders options if the Book Rights Registry allows it to, and the Book Rights Registry will be dominated by author and publisher representatives.

A Guide for the Perplexed Part III: The Amended Settlement Agreement

Back to Jonathan Band, this time in relatively terse explainer mode rather than law journal mode. I’m linking to [a December 18, 2009 feature](#) at *LLRX.com*, a reprint of an earlier publication from ALA, ARL and ACRL. Band describes major changes in GBS2, emphasizing changes relevant to libraries. For those wishing to understand the significance of the amended settlement and lacking the patience for Band’s law review article, [this piece](#) is recommended reading.

His discussion of library issues includes a good explanation of *why* GBS2 excludes most foreign publications and clarifies that Google intended to keep scanning these books. Some other items:

- The new authority of the Book Rights Registry to increase the number of public access terminals in public library branches
- A technical change, clarifying the scanning threshold after which Google can cross-provide digital copies to fully participating libraries—it means 300,000 *volumes*, not titles
- Inclusion of OCLC (or at least non-exclusion) as an institutional consortium for purposes of the agreement
- Clarification that the agreement doesn’t allow for scanning books on microform

- Clarification on privacy, that Google won't provide personally identifiable user information to the Book Rights Registry unless required by a valid legal process
- The new window for rightsholders to request removal of books and what happens to requests after that deadline.

Band also notes rightsholder changes. Among them, I've already mentioned the dominance of authors (but not academic authors) and publishers on the BRR. There's also clearer language on what constitutes an "insert" within a book with separate rights—the "insert" must be separately registered, not just as part of the collected work. I've also already mentioned changes on unclaimed works, but probably not the explicit support for Creative Commons licenses.

Under competitive issues, while Band notes that GBS2 doesn't address the key monopoly issues, he does note broad changes in pricing algorithms for individual books, the explicit inclusion of third-party resellers, the deletion of the "most favored nation" clause, a limit on additional revenue-generating services—and, significantly, the waiver of the *Noerr-Pennington* doctrine, making it possible for antitrust activity to take place even if GBS2 was approved.

If Band showed significant editorial bias in this fine brief discussion, I couldn't spot it—but I suppose that's likely, since I'm in the library arena.

Google Book Search Settlement 2.0: the Latest Scorecard

Now we're into 2010 with Jennifer Howard's [January 29, 2010 article](#) at *The Chronicle of Higher Education's* "Wired Campus" blog. (I'm only including articles from the *Chronicle* and other partially paywalled resources when I, with no affiliation mojo whatsoever, can access them.) The piece appeared a day after the deadline for objections to GBS2 and notes some developments and reactions. Examples:

- Pamela Samuelson and 80 professors wrote Judge Chin about their concerns (specifically Google's monopoly on the digital database)—and Hal Varian at UC circulated a campus response calling the agreement a "huge improvement over the status quo" and saying "it deserves the enthusiastic support of all Berkeley faculty." (A commenter notes that Varian was on leave from UC Berkeley to serve as chief economist for, um, Google.)
- Ursula K. LeGuin (long a copyright maximalist) sent a petition signed by 367 authors opposing the agreement, claiming it allows Google "to circumvent copyright law." (The petition includes a bit of blather about public libraries and "the free and open dissemination of information and of literature"—but only if rightsholders retain full and, presumably, eternal control.)

- James Grimmelmann posted a list of [“Essential Reading for Settlement Junkies,”](#) which I haven’t covered separately and which does point to some interesting items.
- Howard quotes a somewhat typical bit of Kahlian rhetoric from an Open Book Alliance brief, calling GBS2 “more likely a sham and a fraud on the public.”

Some good links to some lively resources, some of them not covered here.

Google Book Search Settlement: Updating the Numbers

This Fred von Lohmann piece at *Deeplinks* appears in two parts—Part I [on February 19, 2010](#), Part 2 [on February 23, 2010](#). These are Google’s numbers; von Lohmann notes that others might dispute some of them. Without the useful discussion—these are brief posts and easy to read—here are the key numbers:

- Total number of books in bibliographic records in the world = 174 million
- Total number of books held by Google partner libraries = 42 million
- Total number of books subject to the amended settlement = 10 million (including those not yet scanned)
- As of February 8, 2010, 44,450 claim forms (that is, forms from 44,450 authors and other rightsholders) and 485 lists had been received, covering 1.13 million books and just under 22,000 “inserts.” Of 1.108 million books claimed online, just under 620,000 are classified by Google as out of print, 488,000 as in print. In other words, rightsholders had claimed about 10% of the works in question.
- Another 6,818 rightsholders explicitly opted out, requesting exclusion, thus representing about 13% of 50,000 rightsholder responses.
- The average claim form (one of those *truly* meaningless averages) is for 895 books, with a relatively small number of publishers claiming most claimed works. In all, 71% of books were claimed by publishers, 29% by authors.
- While Authors Guild claims more than 8,500 authors and AAP claims to represent over 300 publishers (imprints?), 30,000 authors and publishers have already signed up for Google’s Publisher Partner Program

There’s some interesting discussion along with these numbers. As one among the 44,450 (I claimed six books that I knew to be in Google Books and where I had explicit reversion of rights from the publisher), I can attest that the claim process was both well publicized and quite easy. The relatively small number of claims at that point was probably meaningful.

Google Argues for Approval of Book Search Settlement

Norman Oder wrote this news analysis at *Library Journal* [on February 12, 2010](#)—and it not only excerpts some tasty items from Google’s brief (and briefs from the plaintiffs), it includes a Scribd window on the full 77-page Google brief. (All quoted material in this item is from the Google brief.)

I think Oder’s pick for the most striking argument—cited as the subtitle for the *LJ* piece—is Google’s assertion that a monopolistic Institutional Subscription is worthwhile (although the subtitle misses the doublespeak of the excerpt itself):

In sum, granting Google the right to include unclaimed works in the Institutional Subscription serves the pro-competitive goal of making a desirable new product available to libraries, universities and other institutions and has no anticompetitive exclusionary effects on other potential competitors. It is indisputably more pro-competitive and output-enhancing to have one seller rather than none.

It takeschutzpah to assert that a monopoly is pro-competitive. Few would deny that Google haschutzpah. (Oder then links to Robert Darnton’s disagreement.) Nor is Google shy to claim that the new service is not only a library but *the greatest library in history*:

No one seriously disputes that approval of the settlement will open the virtual doors to the greatest library in history, without costing authors a dime they now receive or are likely to receive if the settlement is not approved. Nor does anyone seriously dispute, though few objectors admit, that to deny the settlement will keep those library doors locked while inviting costly, fragmented litigation that could clog dockets around the country for years.

Google points out the hypocrisy of Amazon’s questioning of Google market dominance and offers another mild suggestion that GBS2 would be a Very Good Thing:

Anxieties about what might be best for a particular objector should not become fatal to what is undoubtedly extraordinarily good for all class members and for the general public. The ASA should be approved because it complies with the letter of the relevant laws and advances their purposes beyond measure. The benefits of approval are bounded only by the limits of human creativity and imagination. The costs of disapproval are equally large.

There’s more here, including some comments about GBS and libraries—and I think Oder’s done a good job, so you should [go read it directly](#)

(thus providing eyeballs for lots'o'ads, deservedly). (“ASA”—the Amended Service Agreement—is GBS2.)

Google throws down gauntlet: no more book settlement changes

That’s the headline for John Timmer’s [February 12, 2010 story](#) at *ars technica* and his take on the February 11 filings (Google’s excerpted above)—except that it wasn’t just Google, it was *all* direct parties.

The plaintiffs’ filings largely argue that the ASA meets the needs of the class they represent. As such, their filings focus on the fact that rightsholders will be receiving reasonable payments from Google, and will retain a significant degree of control over the display and sale of the works. In general, these arguments duck the larger legal issues identified by the DoJ and other groups.

Google, in contrast, tackles them head on, but not before reiterating its big-picture take on the settlement: its digitization efforts are the only thing preventing another Library of Alexandria-style tragedy, and making the results available is a public good that should override petty concerns raised by its competitors.

Timmer’s take on Google and antitrust:

Google is also unimpressed by the antitrust worries. Its competitors in the book scanning field, like Microsoft and Yahoo, have dropped out—”There is, in other words, no ‘competition’ to ‘eliminate.’” As for the vending of orphaned works, Google notes that it’s a new entrant to the field, with essentially no market share in books at all. As such, it can’t possibly have monopoly power, and it contends, contrary to the arguments made by others, that it’s unlikely to get it.

There’s more, and it’s another reasonably good take on the situation, although the final paragraph has one troublesome error:

Nevertheless, there are a couple of things that Google could do that would probably get most of its opponents on board: change the agreement to opt-out, and turn its existing digital archives over to a third party. The fact that Google has decided to fight for the existing ASA shows that it’s not interested in either of these solutions, meaning the company definitely wants the rights to orphaned works, and it intends to leverage its digital collection in improving its data analysis capabilities.

Interesting that, two years later, “opt-out” (what GBS *did* call for) still hasn’t changed to “opt-in.”

GBS and the Judgement of Solomon

Here's an interesting opinion piece, [published February 23, 2010](#) on *Exact Editions* by Adam Hodgkin after he read the transcript of the February 2010 fairness hearing. Hodgkin finds himself in “considerable admiration for the American legal system” including the whole *idea* of a fairness hearing. He also admires the process he saw Judge Chin going through—but says Chin “clearly needs” outside help, and suggests the Bible as a guide.

The crucial point is that this is once again a dispute about a child who should have a long and healthy future and there is a danger that it may be smothered or torn apart in his chambers. The orphan books should thrive! But there are too many jealous ‘foster parents’ and the judge will need a masterly stroke if he is to separate the shameful pretenders from the true mother. Is there scope for the judge to put the settlers to a Solomonic test?

After a quote or two from the transcript, Hodgkin suggests that Google’s “with opt-in, there’s no settlement” stance was a bluff “to be called.”

The parties should be forced to live with a purely opt-in solution, which incidentally keeps copyright the right way up, will keep Ursula le Guin, and the French and German governments happy; or (and at this point Judge Chin needs to stroke the handle of his sword, even test the mettle of the blade with his forefinger) Google must be much more generous with the copyrights it has opted from the orphans. Generous to the public domain and non-exclusive to its competitors.

I *guess* the other side of that sword is that Google should be giving away orphan works in their entirety; I may be misreading that. Oddly enough, no comments on this audacious suggestion...but see below.

And that’s it for general overviews and commentary while GBS—the original GBS or GBS2, also known as ASA—was still on the table, although there will be many more topical discussions to come. Now we jump forward a year...to March 2011, when Judge Denny Chin handed down his ruling on GBS2.

General Notes: After the Outcome

You already know the key point: Judge Chin [rejected the proposal](#). This section includes a sampling of commentaries on that decision and what’s happened since, again focusing on overviews rather than specific topics covered later.

Google Books: Copyright Settlement Rejected

That's Kenneth Crews writing [on March 22, 2011](#) on the blog of the Columbia University Libraries Copyright Advisory Office.

To state the decision most succinctly, the court has rejected the proposal, leaving open the opportunity for the parties to renegotiate and resubmit. The case is a copyright infringement claim brought by groups of authors and publishers—as copyright owners—against Google, asserting that the scanning of books and the development of a searchable database is an infringement of copyright. The facts and the litigation are naturally much more complex, but alleged infringement is at the core. The settlement had some important support, but it also encountered significant criticism.

Chin noted that the vast majority of comments received objected to the settlement and found significance in the fact that 6,800 class members had opted out.

In this context, the court examined the somewhat technical question of whether the representative members of the class could adequately represent the many different interests of the multitudes of rightsholders potentially affected by the settlement. Here is where the court came down most bluntly against the settlement. Among the conclusions:

The class representatives would be authorized to establish a registry and a fiduciary to exploit the use of unclaimed books (i.e., orphan works). The court found that Congress, and not the court, is best able to address the interests of orphan works. Moreover, the matter of orphans should not be decided “through an agreement among private, self-interested parties.”

The proposed settlement would give the parties authorizations that go far beyond the original claims raised in the case. One telling statement: “There was no allegation that Google was making full books available online, and the case was not about full access to copyrighted works. The case was about the use of an indexing and searching tool, not the sale of complete copyrighted works.”

The interests of the representatives are sometimes in direct conflict with large numbers of rightsholders. The court mentioned especially that many academic authors do not share the profit motives of the representatives, and the profit motive is at odds with the interests of owners of unclaimed works: “The parties have little incentive to

identify and locate the owners of unclaimed works, as fewer opt-outs will mean more unclaimed works for Google to exploit.”

Chin also objected to opt-out on a fundamental basis and had other issues. He suggested that the parties revise the settlement, clearly guiding them toward an opt-in system.

Crews closes a tight summary with some possibilities as to what might come next:

The parties may accept the invitation to convert the proposal to “opt in,” but that would undercut the ability to include orphan works in the database. The parties could abandon the settlement and return to litigation, but that choice is fraught with expense, delay, and risks. The parties could appeal to the Second Circuit. With so much invested to date, an appeal poses comparatively modest costs and few downsides. The more difficult prediction, however, is whether Congress will take up the court’s challenge and whether it is capable of crafting legislation on this thorny subject that might actually serve the interests of authors, publishers, online services, libraries, and the public.

I can only commend Crews’ successful (and I suspect difficult) attempt to write that last sentence without an emoticon. Oh, I think that’s an *easy* challenge: Congress crafting balanced copyright legislation is a *lot* less likely than Congress adopting single-payer health care.

Federal judge rejects Google book monopoly

Crews tried very hard to offer an objective summary and, I believe, succeeded admirably. Most other writers felt no such compunctions—as in the title of this Timothy B. Lee piece [on March 22, 2011](#) at *ars technica*. A few excerpts:

Judge Chin noted that there were many conflicts of interests between the named plaintiffs (the Authors Guild and the Association of American Publishers) and copyright holders they were supposed to represent. For example, a group of academic authors argued that many academics seek to maximize access to their works, whereas the named plaintiffs were commercial authors and publishers focused on maximizing profits. The settlement was also opposed by numerous groups of foreign authors who argued that their interests had not been adequately represented in negotiations. They also argued that the opt-out requirements were particularly burdensome for foreign authors and that the settlement conflicted with international treaty obligations.

The article refers to “significant antitrust concerns”—a little softer than “monopoly” in the title. Lee does note some objections that did *not* apparently impress the judge, such as arguments that rightsholders lacked adequate notice and privacy concerns.

I’m a little mystified by the last sentence in this paragraph:

It’s unclear where things go from here. The settlement was the product of several years’ negotiation, and Judge Chin took more than a year to hand down his decision. With the online books market evolving rapidly, the case may grow less important as Apple, Amazon, and other competitors build their own digital book empires.

Work being done by Apple and Amazon to scan out of print books and make them visible or available: Zero, as far as I can tell. Is this a technophile’s subtle version of “it’s old, nobody cares anyway”?

Quite a few comments, and some are interesting—but there’s also this from someone who “deal[s] with copyright and IP on a daily basis in our work”:

The concept of orphan works was drummed up by publishers and others who don’t really want to put the time into finding the owners of copyrighted works. They want an out so that they can avoid punitive punishment when they’re caught abusing someone’s copyright.

I’m sorry, but that’s *nonsense*. Publishers don’t give a damn about orphan works; otherwise they wouldn’t *be* orphan works. Later, the pseudonymous commenter offers a similar piece of nonsense:

Orphan works is shorthand for shortcuts to lazy managing of copyrights, and avoiding paying owners of those copyrights, and when they’re caught avoid any penalties possible.

Actually, the final paragraph of the comment is a dead giveaway as to the sort of balanced copyright this person finds proper:

I can imagine an orphan works bill that actually would serve the interests of publishers and authors alike. But what we have now is not it.

Note whose interests are not included in that exhaustive list? Readers, librarians, the public good, the advancement of new creative endeavors. Nope: It’s all about authors and publishers. Period.

Inside Judge Chin’s Opinion

That’s James Grimmelmänn’s title for this fairly long [March 22, 2011 analysis](#) at *The Laboratorium*. (How long? 3,977 words, not including the

comments—less than one-tenth the length of this roundup, but that’s pretty long for a narrow-column blog post!)

I have now had the chance to go through Judge Chin’s opinion rejecting the proposed Google Books settlement for the third time. I am struck by how much—and how little—it says. Its holding is clear and direct: the settlement “would release Google (and others for certain future acts ... an arrangement that exceeds what the court may permit under Rule 23.” (21) The legal analysis supporting this conclusion takes perhaps five pages out of forty-eight. The rest of the opinion is ... well, it’s complicated.

Of seven kinds of objections—notice, representation, future releases, copyright, antitrust, privacy and international—Grimmelmann concludes that “future releases” is the only one found sufficient for rejecting the settlement. Privacy gets short shrift, as does (appropriately, in my opinion) notice. The others are noted but not ruled on.

What is going on here? The future releases issues are sufficient by themselves to reject the settlement; indeed, having concluded that the settlement “exceeds what the court *may permit* under Rule 23” (21, emphasis added), Judge Chin left himself no choice but to reject it. The rest of the opinion provides reasons to support that result—but the opinion is cagey as to which of these are additional legal rationales, and which are just policy arguments.

Why so cagey? Either Chin deliberately wrote a “minimalist opinion” or he rolled everything up into the ultimate question of fairness. I love this sentence: “If hard cases make bad law, then perhaps big cases make strange law.”

The other thing that struck me immediately about the opinion was the remarkable diversity of objectors whose views it cites. It quotes from dozens of different filings, including one notable passage on pages 33 and 34 that pull together concerns from four authors and a pair of literary agents about the settlement’s opt-out structure. This is a quietly effective piece of judicial rhetoric: it emphasizes the range of objectors as well as the range of objections. This goes along with its emphasis on the “great in number” objections and the “extremely high number of” opt-outs: persuading the reader that class members disapprove of the settlement.

That’s the high level—and the first 600 words of the post. The rest looks at details. It’s certainly worth reading. Just a few highlights:

Chin has set up a dichotomy: Google’s past conduct in scanning and searching was the subject of the lawsuit, but it is Google’s

future conduct in selling whole books that would authorized by the settlement. The case “was not about” the same things the settlement is...

Regarding notice, Chin says “Of course, the case has received enormous publicity, and it is hard to imagine that many class members were unaware of the lawsuit” and Grimmelmann notes:

This last phrase reads as though it were directed to Scott Gant, who pushed the notice issue vigorously at the fairness hearing, only to be asked, “You’re here, though?”

But are AG and AAP actually representative of affected parties? Chin directly notes objections from academic authors and foreign rightsholders—but also the orphan rightsholders who couldn’t directly object. As regards whether works *are* actually orphans, the decision notes that the “parties have little incentive to identify and locate the owners of unclaimed works, as fewer opt-outs will mean more unclaimed works for Google to exploit.” As to the typical response—*every* class action represents some people who never come forward—Chin wasn’t entirely buying that in this case. From the decision:

While it is true that in virtually every class action many class members are never heard from, the difference is that in other class actions class members are merely releasing “claims” for damages for purported past aggrievements. In contrast, here class members would be giving up certain property rights in their creative works, and they would be deemed—by their silence—to have granted to Google a license to future use of their copyrighted works. (30)

Grimmelmann loves this: “If you ask me for proof that Judge Chin gets it, I’ll cite this passage.”

There’s a *lot* more here, all of it worth reading. What did Grimmelmann think would happen? His closing thoughts:

If I had to bet, I would guess that we’ll end up with a revised settlement drafted to meet Judge Chin’s specification, which will be approved relatively quickly (at least compared to this last go-round). His opinion is short on specific guidance, but it’s relatively easy to extract the essentials. Here’s what I predict:

Google is allowed to continue scanning and searching in exchange for cash payments on the order of (but perhaps not exactly) the \$60 in the present settlement, and it’s required to provide an opt-out. Very few people have argued that this form of settlement would be beyond the court’s power. The precise explanation of how this would

be distinguishable from the present settlement, although quite feasible, will require some nuance and subtlety.

The Display Uses—Consumer Purchase, Institutional Subscription, etc.—are either gone entirely or are offered on an opt-in basis. The difference between these two possibilities is not large, since, in effect, Google already offers an opt-in through the Partner Program.

The libraries receiving digital copies are released from liability but are even more tightly restricted in the uses they can make than under the present settlement.

The fates of other facets of the settlement such as the Research Corpus, will be hammered out in the negotiations.

My read is that the parties are not enthusiastic about litigation. This has been a long road, they are tired, and the publishing world has moved very quickly from underneath the settlement. They will be happy to have a settlement that lets everyone claim a kind of minor victory, and to be done with the ordeal. A few of the author objectors, who would like to see Google razed to the ground and Mountain View sowed with salt, will continue to object, but most of the others will quietly shuffle away.

And then, the action will shift to Congress. Will Google start putting together a coalition to push for a legislative solution? Who will sign up? What will the proposed compromises look like? Who will oppose it, and with what arguments? And is this the route by which we will get a national digital library?

The Google Books settlement is dead. Long live the digitized book.

I believe there probably are authors who would like to see “Google razed to the ground and Mountain View sowed with salt” (noting that the Googleplex is a tiny portion of Mountain View—roughly *one-third of one percent*, 26 acres out of 12.27 square miles). And what wonderful commentary!

Mostly thoughtful comments, quite a few of them, and I wonder how many other law professors might have felt the same way as Paul Olum:

This is really, really good stuff James. In fact, I daresay that this blog post—put together in about ten hours, full of deep, substantive analysis about one of the more important tech/law opinions of our day—is the high-water mark for law professor blogging. It’s all downhill from here. The rest of us might as well

just start posting cute pictures of kittens from now on, because we can't do better.

Citizen of Google

This odd essay by Jeffrey Pomerantz appeared [March 23, 2011](#) at *PomeRantz*. After admitting that he's not a legal scholar or qualified to be issuing opinions, Pomerantz says:

I think that Judge Chin really blew this one. First of all an opt-in arrangement, as Google has pointed out, is completely untenable. As a result, a vast number of orphan works will be lost for to public use, which is a social tragedy of the highest order. Second, I will grant you that perhaps Google would gain essentially a monopoly over orphan works. However, who else but Google could do this? I don't see Microsoft or Amazon stepping up to this particular plate.

So Judge Chin is wrong *on the law* because Pomerantz liked the desirable outcomes of GBS? It doesn't work that way—and Pomerantz' extended “better a monopoly than nobody” discussion (I only include a bit of it) is not convincing. Nor am I thrilled about this: “I say this to my classes all the time, and I'm sure my students are tired of hearing it, but Google is fighting libraries' fights for us, and has been for years.” Nope. Google fights nobody's fights but its own; otherwise, it would have pursued the fair use defense. Actually, Google's convinced me in the past that it will buddy up to libraries just as long as it thinks it has something to gain, then pretend that it never heard of them.

After discussing a separate issue, Pomerantz offers a breathless love letter to Google and I find that after quoting a paragraph, I couldn't bear leaving it in. You'll have to follow the link. Even Apple rarely gets this level of adoration. Screw the laws: Google does things Pomerantz wants done, so he'd like to be a citizen of Google. For all its faults, I'll stick with the U.S.

Good and Bad in Google Book Search Settlement Decision

This post, by Corynne McSherry [on March 23, 2011](#) at *Deeplinks*, is a surprisingly different version of “we know copyright law better than Judge Chin,” this time from EFF's perspective. Surprising in part because of a claim that Chin “acknowledged the importance of the privacy concerns we helped to raise”—which other observers seemed to see as a handwave. Also, to be sure, that the court agreed with an EFF board member that academic authors might not share the interests of the Authors Guild.

But, the post says, “the court also got some things fundamentally wrong in its copyright analysis.” Namely, the general right of a copyright owner to prevent publications, and finding that it's unreasonable to ask

rightsholders to opt out. Oh, and paying attention to those foreign rightsholders.

A Copyright Expert Who Spoke Up for Academic Authors Offers Insights on the Google Books Ruling

That lengthy title appears over Marc Perry's [March 23, 2011 interview](#) of Pamela Samuelson in *The Chronicle of Higher Education*. Professor Samuelson advocated for academic authors as *not* sharing the interests of the Authors Guild—and Judge Chin did raise the issue of whether AG adequately represented the interests of all authors. A few excerpts (all direct quotes from Samuelson):

It's the only ruling really that the judge, I think, could have made. The settlement was so complex, and it was so far-reaching. With the Department of Justice and the governments of France and Germany stridently opposed to the settlement, it seems to me that the judge really didn't have all that much choice. So the ultimate ruling, that the settlement is not fair, reasonable, and adequate to the class, is one that I think was inevitable...

Academic authors, on average, would prefer open access. Whereas the guild and its members, understandably, want to do profit maximization.

I would love to believe that most academic authors “would prefer open access,” but the troubled history of OA doesn't confirm that.

I think this comment is the best response to those who bemoan the ruling because GBS could have done so many good things:

Many of the things that the settlement would do are copyright reforms that I think are good. The question is, Can you do this through a class-action settlement? One of the things that was very pleasing to me about the judge's ruling is that the judge also said changes this far-reaching to the default rules of copyright law have to be done through Congress.

The settlement would grant Google about five different licenses that ordinarily, to get that broad a license, you'd have to get it from Congress. It's a license to scan all the books and to store them. A license to make nondisplay uses of them for purposes such as improving search technologies and automated translation tools. It would grant a license for nonprofit researchers to engage in “nonconsumptive” uses—so research uses for academic purposes. It would grant Google a license to give “library digital copies” of the books scanned from library collections back to those libraries and

allow the libraries to make certain kinds of uses of the works. And it would give Google a license to commercialize all of the out-of-print books in the corpus. It's really quite extensive.

If Congress was going to grant licenses like this, it wouldn't just grant them to Google. Part of what the Justice Department came to recognize is that the licenses that Google would get from the settlement would create barriers to entry to any other firm, because no one else could get those licenses. That's something that really fed into the antitrust analysis in the case. The settlement would give Google a *de facto* monopoly over the orphan books [unclaimed works whose copyright owners aren't known or can't be found] that would make a subscription service that it could offer unreachable by any subscription service that anyone else might offer. Google could have millions and millions of books that no one else could reach.

Samuelson also discusses privacy issues (which Chin noted but didn't find sufficient to reject the settlement) and the clear fact that Google *lacks* library attitudes on reader privacy and was unwilling to make appropriate commitments. "Trust us; we're not evil" doesn't do it.

A section of the interview has Samuelson speaking as the voice of academic authors—and, frankly, I'm no more satisfied with that than I am with AG speaking for all authors. Is there really a unified class of academic authors with common interests? Take, for instance, this:

One path is that academic authors can communicate with Google about their interest in making their books available on an open-access basis. That would be something that would allow more of their books to be more widely available.

I can't prove that it's *not* the case that most academic authors would be enthusiastic about this idea, but I'm skeptical. Her second path is working with "a group of academics" to put together a legislative package—and third, there's litigation, where "I think academic authors will probably offer support to Google in its fair-use defense, because we are the kind of people who think that if you scan my book in order to index it and make little snippets available, that's actually a good thing."

Here's the first paragraph of Samuelson's four-paragraph comment on the prospects for legislative change, and it sounds as complicated as I'd expect:

It would require a lot of energy, and a lot of coalition building. But I think that there's some possibility of it, actually. I'm not wildly optimistic about it. There is this amazing vision of access to knowledge that a lot of people are in favor of. If that's true, then we

ought to be able to come up with something that would make it all work.

She does say “All of the major parties have been in favor of orphan-works legislation.” I wonder what that means—what an orphan-works legislative approach favored by AG and AAP would actually look like.

There’s quite a bit more in what’s an interesting set of perspectives.

GBS Update: The Settlement Is Dead; Long Live the Settlement Negotiations!

That’s Charlie Petit posting [on March 23, 2011](#) at *Scrivener’s Error*. Petit, a lawyer who focuses on publishing issues from a pro-author perspective and who believes in moral rights for authors and other content creators, precedes this essay with links to a careful (and snarky) [essay on the suit itself](#) (a “curse on both sides” essay) and another essay [from October 2008](#) taking the proposed settlement to pieces. (I’d call it a fisking, and that’s the tone, but you can’t fisk a 300-page proposal in even a very long online article.)

Here he does the same for the rejection, but relatively briefly—and while Petit points to Grimmelmann for extensive commentary, he thinks “he missed some of the civil-procedure-type nuances.” (He also points to two other writers.) The following excerpt may give a sense of Petit’s calm tone while staying within fair use (his sidebar suggests a litigious nature regarding reuse of his material):

I’m not going to cover the various blatherings of the putative parties to the settlement; neither am I going to cover the loons (and you know who you are; but just because you’re not on the list below doesn’t mean I think you’re a loon).

He does not regard the decision as “a model of clarity” and thinks Grimmelmann’s conclusion, while likely to become the majority perception, is shortsighted because it ignores procedural issues in favor of policy ones. He believes antitrust will be important in future proceedings. He did *not* believe an immediate appeal would succeed (“slightly (but only slightly) more than a snowball’s chance in hell”), a good call. He expected a return to the bargaining table, “with the Authors Guild still trying to shut out all other authors’ groups.”

All in all, an interesting and very different set of informed perspectives from a practicing attorney specializing in this area.

Please Refine Your Search Terms

This Steve Kolowich report, [appearing March 23, 2011](#) at *Inside Higher Ed*, notes the rejection of the settlement and quotes a number of commentators. I question Kolowich’s definition of orphan works as “books for which there is no clear copyright holder”; rather, they are works for

which the rightsholder *can't be contacted*, which is a different thing entirely.

Reading through the notes and comments, I become aware again that supporters tended to focus on the possible good outcomes while largely ignoring the question of whether the outcomes represented a fair settlement of the suit. Jeanine Varner of Abilene Christian is doubtless correct in saying the settlement “is a significant change for the better by creating a means for us to offer immediate electronic access to crucial published resources”—but it might still be bad law. Kolowich calls the decision “light on references to libraries, students, and research”—which makes sense, given that libraries, students and research *were not parties in the suits or settlement*.

The Book Deal May Be Dead, But Google Is Still Right

No waffling on Mathew Ingram's part in [this March 25, 2011](#) *gigaom* item—but he's not really referring to GBS itself, but to the original issue:

But the fact that the arrangement has been rejected might not be such a bad thing, because it puts the spotlight back where it should be: on the fact that Google is doing nothing wrong—legally or morally—in scanning books without the permission of the authors or the publishers of those books.

Ingram calls the plaintiffs' stance “ridiculous” and goes back to fair use. He notes the monopoly issue and calls it “arguably over-reaching” but concludes:

But that doesn't change the fact that Google's initial impulse was the right one: it does have the right to scan and display extracts from books, regardless of what the Authors Guild and the AAP say, and it should continue doing so.

I wonder whether Google *will find the spine to defend fair use in this context. Some of us found GBS disappointing because Google was caving on fair use; it now has a second chance—albeit one that doesn't let it set up a profitable secondary enterprise.

A small set of comments, some reasonable—but also one from a reader who believes that the overreaching statements that appear on copyright pages *must* be part of copyright, and therefore that even Google's scanning must be infringement.

To the Whingers Go the Spoils in the Google Books Decision

This Ryan Singel post, [on March 29, 2011](#) at *Wired.com*, is one of the more mean-spirited commentaries on the decision, from someone who apparently knows the law much better than, say, James Grimmelman. Maybe the title's enough, along with Singel's assertion that “the world will be poorer for the decision.” He calls anybody who objected “the

copyright whingers,” specifically snarks at authors mentioned in the decision and says:

Yes, the paranoid and the curmudgeonly get the veto over the library of the future because, well, it might actually get them readers.

As I find a one-fingered salute rising unbidden, I note that the writer has no interest in the *actual* reasons for the decision. Nor does he have any doubt whatsoever regarding the outcome if Google had defended itself in court: “The authors would have lost in court.” “It’s very clearly fair use in the United States for Google to digitize any copyrighted book and use snippets of it in search results.” No question, no doubt: This was 100% fair use, absolutely guaranteed.

If Google had fought this suit on those grounds, as many digital rights groups hoped it would, it would have likely won and set a precedent for other innovators who often find themselves crushed by lawsuits from organizations like the MPAA and RIAA.

Then Singel misstates the settlement at least in one regard: “The settlement provided hundreds of dollars each to authors whose books had been scanned.” No, it did not. I signed up for the registry. There was never an offer of “hundreds of dollars each”: \$60 is not “hundreds.”

We learn that Chin was really punishing Google for being innovative, which is an interesting read.

So here we have it. Google was naughty for not asking permission from every schmuck in the world who owns a copyright, before it dared to try to create the library of the future. A library that would let anyone with a net connection—rich, poor, blind and sighted alike—search, sample, read and buy nearly any book ever published (at least those published in the United States).

As one of those schmucks, I find this wording deliberately and needlessly hostile. I’m also interested in the extent to which Singel faults Chin and dismisses the rights of authors (oh, sorry, “schmucks.”) Singel recognizes that Congress probably won’t pass orphan works legislation—and seems to conclude that this makes Judge Chin not only wrong but a tool of the copyright maximalists. A sad piece of work, albeit what I’d expect from *Wired* (unless, of course, Condé Nast’s intellectual property is at issue).

Google Book Settlement Rejection: A Missed Opportunity

Bill Rosenblatt, writing [on March 30, 2011](#) at *Copyright and Technology*, was also unhappy about the decision, but didn’t feel the need to be a jerk about it. This is a calm and fairly subtle discussion mostly related to what Rosenblatt sees as a failed opportunity to establish the Book Rights Registry *as an industry tool*.

The interesting thing is that Rosenblatt seems to be more of a copyright maximalist—noting that “large commercial entities” use lawsuits because they can’t get legislatures to do their bidding rapidly enough. Those lawsuits are almost uniformly intended to tighten copyright restrictions, not broaden user rights.

Neither is Rosenblatt focused on increased fair use or anything of the sort. No, he wants the BRR because it would “improve the global copyright scene for the digital age”—and “improve” pretty clearly means “for business.”

Many of the problems in managing digital rights to content could be solved if there were complete, consistent, up-to-date, and easily accessible sources of information about content and rights holders. Private companies have made various attempts to solve this problem over the years; none have succeeded, owing to unrealistic profitability requirements, overly narrow scope, lack of cooperation from rights holders, and other factors.

And Rosenblatt thinks BRR should include *everything*—on an opt-out basis.

Now, with Judge Chin’s rejection of the settlement, the BRR looks like a lost cause. Judge Chin’s opinion suggests that a revised settlement could be approved if it works on the “opt in” instead of “opt out” principle, i.e., it should include only those works whose copyright owners proactively agree to let be included. This may pass various legal sniff tests. But any resulting Book Rights Registry under an opt-in regime would be of highly dubious value to the industry in general; in fact, it would scarcely differ from repositories of licensable material available today, such as Overdrive’s Content Reserve.

This discussion may be orthogonal to most others: it’s not about improving citizen or library access; it’s about making it easier to license material.

Six Reasons Google Books Failed

Robert Darnton published this [on March 28, 2011](#) in the NYRblog from the *New York Review of Books*. A longer version appeared in the April 28 print edition. It wasn’t until I skimmed down to the comments that I realized my mind had added a word Darnton leaves out of the title: “Settlement.” Has Google Books actually failed? That seems a bit sweeping.

Dealing with the *actual* failure, Darnton sees, well, six “crucial points where things went awry”:

- He says “Google abandoned its original plan to digitize books in order to provide online searching”—but *that’s not true*. GBS would have

expanded that plan to involve other services, but certainly not abandoned digitizing and searching.

- I don't see a clearly identified second point, since the second paragraph expands on the first point.
- "Third, in setting terms for the digitization of orphan books—copyrighted works whose rights holders are not known—the settlement eliminated the possibility of competition." GBS2 covered that, but it's certainly true that GBS2 still "amounted to changing copyright law by litigation instead of legislation."
- Fourth is the foreign rightsholder issue—again largely covered by GBS2.
- "Fifth, the settlement was an attempt to resolve a class action suit, but the plaintiffs did not adequately represent the class to which they belonged." Absolutely true.
- "Sixth, in the course of administering its sales, both of individual books and of access to its data base by means of institutional subscriptions, Google might abuse readers' privacy by accumulating information about their behavior." Also true enough.

Again, these are all reasons for failure of GBS, not Google Books. Darnton then enumerates some of the good that could have come from GBS—and, sigh, says "these advantages can be preserved without the accompanying drawbacks" by creating a Digital Public Library of America, the seriously misnamed proposal that Darnton's heavily involved in.

The rest of the piece is largely about DPLA and what Darnton sees as similar European initiatives. I'm not dealing with DPLA here (and possibly not anywhere), so I'll refer you to the original essay—which, it turns out, is just as misnamed as DPLA.

Google Books Settlement, 2008-2011

This first of three general commentaries from later in 2011, after the dust had settled. This one's by James Grimmelmann, [posted August 17, 2011](#) on *The Laboratorium*—and it's an obituary of sorts, as the title suggests.

The Google Books settlement, a book collector whose audacious plan to remake copyright law was ultimately for naught, died today. It was caught in the blast from a recent court decision, and received fatal injuries. Ironically, the settlement, which had been seriously injured in the spring, had been rumored to be planning a comeback tour. In the end, however, doctors declared that its internal divisions were incurable. The settlement was a little over two months short of its third birthday, and is survived by millions of orphan works.

The fatal blow, discussed in some detail, is a decision on a “kind of older sibling to the Google Books case” in which freelance writers sued databases for including articles without authorization—and that case has been around for a *long* time, going to the Supreme Court in 2001 and 2010.

Most recently, it has had the form of a proposed class-action settlement on behalf of all the freelancers that would have paid them up to \$18 million in exchange for letting the databases reproduce the articles in perpetuity. The Second Circuit held that the settlement couldn’t be approved because different parts of the class were so at odds with each other that they each needed their own lawyers in the negotiations. Since the deal was worked out by a single group of lawyers for the whole class, that obviously hadn’t happened, and it’s back to square one for the settlement talks in the freelancer lawsuit, or maybe even square zero if some of the parties give up on settlement entirely.

Grimmelmann spells out some of the parallels, but I went “Bingo!” as I read that paragraph: Neither AG nor AAP represents a uniform class with identical interests. There’s more discussion and it’s interesting, but it relates to the other case, not GBS.

For GBS, Grimmelmann concludes, “square one” isn’t an option: There are too many subclasses of authors and publishers for a plausible and fair settlement to be reached.

Creating subclasses that track these different groups, then supplying them with their own skilled lawyers, and putting everyone in a room together to knock out a new settlement: it just isn’t going to happen. The Venn diagram will have at least a dozen different boxes in it. The expense would be absurd, it would take months or more likely years to pull off, and I still can’t imagine those negotiations succeeding, especially not after the level of vehement opposition to the original settlement. There is no trust here, and Judge Chin had already been banging heads together to get the case moving. No, the Google Books settlement—*any* settlement—is now dead. There is no square one: this case is going back to litigation.

There’s more here (and discussion of yet another class action lawsuit involving Google), but this is the key section relating to Google Books.

One Google Books To Rule Them All?

Maria Bustillos, writing [on October 26, 2011](#) at *The Awl*, opens lively:

Hellzapoppin’ in the world of intellectual property rights these days. Lawsuits, corporate flim-flamming, the claims of far-sighted

academics and developers, furious authors and artists and the conflicting demands of a sprawling Internet culture have created a gargantuan, multi-directional tug-of-war that will inevitably affect what and how we will be able to read online in the future. Recent developments indicate, amazingly, that there are grounds for hope that the public will in time benefit from the results of this epic tussle.

What are those grounds? After a discursion on how Sonny Bono and others have kept things under copyright for ridiculously extended periods, Bustillos uses the subheading “Scan and Be Damned” for what she asserts Google did—and note here a truly unfortunate set of scare quotes:

Copyright law being the morass that it is, Google was in something of a bind when it set out to create the Google Books project. How would they get round the proscription against reproducing books still in copyright? Their solution was just to scan the hell out of everything, make certain results available only as “snippets,” and claim this practice as “fair use.” As even the government’s own guidance on this policy points out, “The distinction between fair use and infringement may be unclear and not easily defined.”

Yes, she consistently scare-quotes fair use and never mentions that it’s *part of the law*. But never mind. Noting that Google wants to keep adding to its index because that means you’ll see its ads more often, she asserts:

This latter point is beyond infuriating to authors, who have enough trouble trying to keep body and soul together as it is. Already Google is profiting from their work, not by selling unauthorized copies of it, which would be illegal, but by selling advertising alongside bits of it. This reasoning was bound to be challenged in court, though many copyright experts thought the “fair use” argument still might fly.

Heck, those legal experts at *Wired* thought it was dead certain, although actual, you know, lawyers weren’t quite so sure. Anyway, the piece goes on to the settlement, its rejection and—oddly—the note that “The settlement negotiations continue, with a new hearing scheduled for next spring.” I do give Bustillos credit for asking Ursula Le Guin to say what she thought an ideal outcome of the Google Books case would be (Le Guin *quit the Authors Guild* because they negotiated a settlement) and got this striking answer:

“Their agreement, or a ruling requiring them, to immediately stop digitalising copyrighted books without obtaining permission from the copyright owner. With a reminder to the libraries that have

been facilitating this illegal activity that it is piracy, and they should not have agreed to it.”

Whew. Scanning is *piracy* and take that, you evil libraries.

I'm not ready to buy Bustillos' assertion that Google and Amazon have an effective duopoly with *no serious competition* for textual information—"for information that would before have been contained in books," a wording that suggests to me that Bustillos is a digital triumphalist who regards print books as already dead. But no, it's apparently *already the case* that the only ways an author can reach an audience are through Amazon or Google. Who knew?

Here's another passage that, while it's not directly related to GBS, gives me more than a little pause:

The job of the modern scholar or critic is to read widely and bring his findings to a public of interested fellow-seekers, something more like a DJ for culture.

This is a long posting and I'm not dealing with all of it. Bustillos is enthusiastic about DPLA as a solution; that's her good news. Color me unconvinced—and wondering why the scare quotes and silly title.

The Elephantine Google Books Settlement

That's James Grimmelmänn again with the final item for this section, a December 5, 2011 post at *The Laboratorium* that's really mostly pointing to [his article of the same name](#), appearing in the *Journal of the Copyright Society of the USA*. At 24 pages (including footnotes), that article is short enough to be worth reading for many of you—and I'll leave you to it, quoting the abstract here.

The genius—some would say the evil genius—of the proposed Google Books settlement was the way it fuses legal categories. The settlement raised important class action, copyright, and antitrust issues, among others. But just as an elephant is not merely a trunk plus legs plus a tail, the settlement was more than the sum of the individual issues it raised. These “issues” were really just different ways of describing a single, overriding issue of law and policy—a new way to concentrate an intellectual property industry.

In this essay, I argue for the critical importance of seeing the settlement all at once, rather than as a list of independent legal issues. After a brief overview of the settlement and its history (Part I), I describe some of the more significant issues raised by objectors to the settlement, focusing on the trio of class action, copyright, and antitrust law (Part II). The settlement's proponents responded with colorable defenses to every one of these objections. My point

in this Part is not to enter these important debates on one side or the other, but rather to show that the hunt to characterize the settlement has ranged far and wide across the legal landscape.

Truly pinning down the settlement, however, requires tracing the connections between these different legal areas. I argue (Part III) that the central truth of the settlement is that it used an opt-out class action to bind copyright owners (including the owners of orphan works) to future uses of their books by a single defendant. This statement fuses class action, copyright, and antitrust concerns, as well as a few others. It shows that the settlement was, at heart, a vast concentration of power in Google's hands, for good or for ill. The settlement was a classcopyrightsliphant, and we must strive to see it all at once, in its entirety, in all its majestic and terrifying glory.

That's it for the overviews. Now, on to pieces with more limited focus (or at least that's how I chose to classify them), including—a bit later on—a whole *bunch* of material related to libraries and metadata.

Orphan Works

I'm probably mischaracterizing one or two of these items, some of which cover much more than orphan works—but that's the facet that struck me as particularly interesting.

The Google Book Search Settlement: Ends, Means, and the Future of Books

Take this one, for example: a 17-page PDF [published April 2009](#) by James Grimmelman for the American Constitution Society for Law and Policy. Do note the date: At this point, the original GBS was under consideration.

It is difficult to overstate the importance of this settlement. The ongoing shift to electronic publishing is arguably the biggest transformation in books since Gutenberg's invention of the printing press. The scale of Google's plans boggles the mind. If the settlement is approved, Google will have the closest thing to a universal library the world has ever seen. We should be enthusiastic about the prospect of creating such a library, and concerned that it may be under the exclusive control of one company. This issue brief will connect this enthusiasm and this concern to the structure of the settlement that gives rise to them both.

While the first part of the brief analyzes the lawsuit and settlement in general, the second part focuses on “one especially important part of the public interest context,” orphan works. He regards that part of GBS as “good for the public to the extent that it makes [orphan works] available again, but potentially bad to the extent it turns Google into a dominant platform with control over a huge catalog of books that no one else has access to.” The third part deals with process—and Grimmelmnn believes orphan works issues need to be resolved through legislation: “Laundering orphan works legislation through a class action lawsuit is both a brilliant response to legislative inaction and a dangerous use of the judicial power.”

The whole brief is worth reading. Grimmelmnn always had mixed feelings about Google’s fair use claim: He recognized that success would usefully increase use of fair use but wasn’t convinced that it was a slam-dunk. He also gets copyright right in his introduction to the second part (emphasis added):

Copyright is designed to increase the supply of creative works available to the public. To do that, it gives creators incentive to create new works by giving them a revenue source; willing buyers pay for copies of the work. Under ordinary circumstances, a user is more than happy to pay a price the owner is more than happy to accept.

An “orphan” work, however, has (or might have) an owner who cannot be found, who may not even know that she is a copyright owner. For instance, think of an author who dies without a will. Her next of kin may have no idea that they are now copyright owners. Or think of a publishing house that gets into financial distress and has to sell itself to a liquidator; the buyer may be thinking of the presses and the office chairs, not the copyrights. Especially with older works that are not currently generating revenue, it becomes all too easy to lose track of ownership records.

If only more lawyers and other writers commenting on copyright issue would begin with a sensible paraphrase of the Constitutional clause! A good, crisp discussion of why orphan works represent a lose-lose situation follows. Then he turns to GBS itself.

It’s important to recognize the critical role that the settlement’s treatment of orphan works plays in its ambitious scope. Because Google is allowed to presume consent of absent copyright owners—precisely the presumption that the plaintiffs objected to when they filed the lawsuit—it sets a default that most of the books in existence in the United States will be part of Google’s collection. Every orphaned

book is a book whose owner will never reverse the default, will never opt out. Google's book services will be comprehensive on a previously unimaginable scale; the settlement ensures that copyright claims by orphan works owners will not threaten that comprehensiveness.

But for all that goodness, "the devil, however, is in the details." The deal was *only* good for Google; its nature as a class action settlement created a huge barrier to entry for any potential competitor; and more.

Finally, there's the issue of process. Grimmelmann's no supporter of the current copyright regime: The first sentence of section III is "Copyright law is broken, and the orphan works problem illustrates everything wrong with it." But it's a legislative problem and he believes it should stay that way.

While some of Grimmelmann's points have been mooted by more recent events, the brief is still worthwhile, powerful reading, ending with this conclusion:

The Google Book Search settlement serves respectable ends through questionable means. The copyright interests in books have been scattered to the four winds over the years, harming both the reading public and copyright owners themselves. True, a class action is a device for gathering together lots of widely scattered interests, but in this case, it's the wrong device. Because this deal was struck through private negotiation among a few parties, it neglects the broader public interest in some critically important ways.

The need for change is real, but at the same time, it's reassuring how eminently solvable the problems with the settlement are. The settlement may have emerged from a questionable bargaining process, but the end product bears at least a familial resemblance to an agreement of which we could all feel proud. This settlement does not need to be problematic, and we should not let it be. The court is being asked to place its imprimatur—our imprimatur—on this reshaping of our copyright law and our publishing system. We the people have the right to insist that our interest, the public interest, be reflected in the outcome.

Legally Speaking: The Dead Souls of the Google Booksearch Settlement

This [April 17, 2009 post](#) by Pamela Samuelson on *O'Reilly Radar* also appeared in the July 2009 *Communications of the ACM*. The quick version of Samuelson's focus:

This column argues that the proposed settlement of this lawsuit is a privately negotiated compulsory license primarily designed to

monetize millions of orphan works. It will benefit Google and certain authors and publishers, but it is questionable whether the authors of most books in the corpus (the “dead souls” to which the title refers) would agree that the settling authors and publishers will truly represent their interests when setting terms for access to the Book Search corpus.

She describes the portions of GBS that relate to orphan works and how a class action suit could result in a license that affects millions of rightsholders not party to the suit—and the extent to which GBS would create a Google monopoly. The “Dead Souls” section makes a nice play on words, connecting Nikolai Gogol’s story “Dead Souls” to Google’s “dead souls” scheme. Going beyond that, Samuelson notes—correctly—that the BRR governing board would almost certainly be dominated by copyright maximalists, who might not really represent the wishes of orphan works authors, especially scholarly ones.

If asked, the authors of orphan books in major research libraries might well prefer for their books to be available under Creative Commons licenses or put in the public domain so that fellow researchers could have greater access to them. The BRR will have an institutional bias against encouraging this or considering what terms of access most authors of books in the corpus would want.

While GBS2 was better in this regard, it’s still a valid objection. Samuelson concludes GBS would bring about greater access to books collected by major research libraries—but at too high a price, two complementary monopolies. She concludes:

The Book Search agreement is not really a settlement of a dispute over whether scanning books to index them is fair use. It is a major restructuring of the book industry’s future without meaningful government oversight. The market for digitized orphan books could be competitive, but will not be if this settlement is approved as is.

An interesting and sometimes wild range of comments, with one writer asserting that the purpose of copyright is “to protect the intellectual property of authors” and labeling libraries “the worst thief one could think of” and another flatly denying that orphan works exist. Pamela Samuelson made a diligent effort to respond (calmly and thoughtfully) to nearly all of the comments.

Google Book Settlement, orphan works, and foreign works

This discussion was posted by Peter Hirtle [on April 21, 2009](#) at *LibraryLaw Blog*. Hirtle thinks the focus on orphan works is ignoring

what he calls “the real losers in the settlement: the thousands of foreign authors whose books can be exploited with impunity by Google and the Books Rights Registry.” He believes most foreign rightsholders for out-of-print books will fail to register with the BRR, despite Google’s publicity efforts.

There are a couple of reasons for this. For one, they may not know that their book is still protected by copyright in the US. In addition, they may assume that international network of reproduction rights organizations would manage their royalties, and not understand the need to register separately.

He’s not sure how big a group that is, but seems to suggest it could be in the millions, and concludes “If there is an injustice being done in the settlement, it is with foreign authors.” I’d guess GBS2 took care of most of that issue, since it *explicitly* excludes foreign works not explicitly registered in the U.S. (except works from Canada, the UK and Australia).

Why the Google Books Settlement is better than orphan works legislation

Another *LibraryLaw Blog* post by Peter Hirtle, [this one dated May 27, 2009](#), and taking a different tack than James Grimmelmann. Hirtle says that books from “inactive rights holders” in the Book Rights Registry aren’t necessarily *orphan* works—that they include works by rightsholders “who could be easily located but who have chosen not to sign up with the Registry.” (Here again, his focus is foreign authors.)

Hirtle then runs through some numbers to try to estimate the number of orphan works. It’s a tricky process—for example, he uses *Global Books in Print* as a probably-too-high outer limit, but *Books in Print* omits a growing number of books that do not have ISBNs. This parenthetical comment on books in copyright but out of print shows just how tough this can be; it’s quoted exactly as it appeared:

(Some of these would be American works that have not had their copyright renewed and hence are in the public domain, but I think the number could only be ~~150,000~~ 1.7 million at most, ~~and so I am going to ignore that~~).

An update paragraph explains that striking strikeout, which changes by *an order of magnitude* one piece of the puzzle. Running more numbers, Hirtle concludes that there might be about 1.4 million *true* orphan works—and another 10.6 million “would either have rights holders who registered with Google or who choose not to register.” That leads us to the two paragraphs that Hirtle believes justify the post’s title:

Even with orphan works legislation, these works would not be eligible for inclusion in a digitized books database since they are

not true orphans. The Google Books settlement is the only way to get cost-effective access to them.

What we need in the settlement is a compulsory license that would allow anyone to license the use of a work maintained by a non-active rights holders, and not just orphan works.

The problem, to be sure, is that this compulsory license would be a judicial fiat, not a legislative act—and that’s problematic, especially since those 10.6 million works are by authors pretty clearly not well represented by the parties in the settlement.

Google Books, and missing the opportunities you don’t see

John Mark Ockerbloom at *Everybody’s Libraries* [on September 15, 2009](#)—and this time he was beginning to think “there was a significant likelihood that the settlement might fall apart,” which he had not previously believed.

There are a number of people in different communities, including libraries, who hope this happens. I’m not one of them. I’m not a lawyer, so I can’t comment with authority on whether the settlement is sound law. But I’m quite confident that it advances good policy. In particular, it’s one of the best feasible opportunities to bring a near-comprehensive view of the knowledge and culture of the 20th and early 21st centuries into widespread use. And I worry that, should the settlement break down, we will not have another opportunity like it any time soon. The settlement has flaws, like the Google Books Project itself has, but at the same time, like Google Books itself, the deal the settlement offers is incredibly useful to readers, while also giving writers new opportunities to revive, and be paid for, their out of print work.

There’s the rub: Can you actually get good policy through bad law, especially bad law that’s the result of judicial rather than legislative action?

Ockerbloom’s description of the problem is good, although much of it (discoverability) only requires indexing and snippets, not the vastly expanded scope of GBS. He also believes GBS might *encourage* Congress to do something about orphan works and “general copyright reform” as a “compelling example.” I wonder.

580,388 Orphan Works—Give or Take

Michael Cairns makes that claim in the title of [this September 9, 2009](#) post at *Personanondata*. He says “no one has attempted to define how many orphan works there really are” (see Hirtle above for just such an attempted

definition, but never mind) then offers this absurdly precise estimate. How does he get there?

Well, I admit, I do my share of guess work to get to this estimate, but I believe my analysis is based on key facts from which I have extrapolated a conclusion. Interestingly, I completed this analysis starting from two very different points and the first results were separated by only 3,000 works (before I made some minor adjustments).

Then it gets strange, when Cairns accuses others of “deliberate obfuscation and lazy reporting” for saying “millions” of orphan titles—except that, if you accept Cairns’ definition of work vs. title, he admits that such reporting may be correct. But, again, never mind.

How does Cairns arrive at his precise number? His first method uses *Books in Print* (notoriously incomplete) and Worldcat. His second uses Bowker’s annual industry data reports—which are almost certainly incomplete. (And, hmm, he’s using both Bowker and Bowker to arrive at his numbers.) Now we get the real precision work, after he’s made the apparent assumption that percentages in Worldcat will fit the Google Books universe:

In order to complete the analysis to determine a specific orphan population, I reduced my raw results based on best guess estimates for non-books in the count, public domain titles and titles where the copyright status is known. These final calculations result in a potential orphan population of 600,000 works. I also stress-tested this calculation by manipulating my percentages resulting in a possible universe of 1.6mm orphan works. This latter estimate is (in my view) illogical as I will show in my second analysis.

So it’s a reduction of questionable percentages based on *best guesses*. Sounds precise to me. His second method used his set of guesses applied to publishing title counts, with some other suppositions added in.

He draws lots of conclusions from his precise analysis, but given my opinion of the analysis I choose not to discuss them. This seems mostly an attempt to beat down any sense that orphan works would result in significant revenue for anybody involved. (Cairns has mostly been a publishing consultant.)

Advantage Google

This essay by Lewis Hyde was published in the *New York Times Sunday Book Review* [on October 1, 2009](#). It’s primarily devoted to orphan works as part of GBS—and Hyde does use “millions” based on relatively direct evidence (rather than Cairns’ hocus-pocus): He believes there are between four and five million orphan works among those Google had

already scanned at that point. His comment on GBS as an orphan works solution:

This is a smart way to untangle the orphan works mess, but it has some serious problems, the most obvious being that it treats orphans as if they were Brats who can be set to work for families who had no hand in their creation. Nothing in the history of copyright can possibly allow for such indenture. In an essay written late in life, James Madison explained that copyright is best viewed as “a compensation for a benefit actually gained to the community.” There were good reasons, he wrote, to give authors a “temporary monopoly” over their work, “but it ought to be temporary” because the long-term goal is to enrich public knowledge, not private persons.

(You need to read the article to understand the “Brat” reference properly.) He believes an independent guardian makes more sense and that GBS would establish monopoly power over orphan works for Google. The piece is well written and you may find it worth reading.

GBS: A Legislative Solution?

Back to *LibraryLaw Blog* and Peter Hirtle, this time with a fairly long post [on November 8, 2009](#). This time, Hirtle really is talking about orphan works—and, after noting various opinions on several sides of the issue (including the cockeyed view of Brewster Kahle that *all* out-of-print works should be regarded as orphan works and that all such works should be wholly available for noncommercial copying unless an author can *prove* ownership, “with penalties for overreaching”), Hirtle concludes that legislative solutions are either unlikely or wouldn’t really solve the problem. Therefore, he believes, GBS should be adopted.

Gripes over Google Books go technical

Let’s finish this section with Larry Downes’ [February 11, 2010 piece](#) at *CNet News*—and it may be worth noting that “go technical” here clearly is not intended to mean “become hyperimportant” but rather “look like [trivial] technicalities.”

Downes looks at the Department of Justice’s objections to GBS2 and says most of them are now addressed to “the manner in which the deal has been constructed—specifically, the use of class action litigation to break the legal logjam of U.S. copyright law.” He notes that DoJ also notes antitrust concerns, although Downes uses scare quotes around “concerns.” (Really? There are no *legitimate* monopoly issues? See the next section.)

It's probably fair to say, as Downes does, that "the government itself" caused most of the orphan works problem through the steady extension of copyright terms and dropping registration as a requirement.

The ASA would largely solve the orphan works problem, for which the government believes the parties "should be commended." The Justice Department, however, still won't endorse the solution. Its particular objection now is the use of the class action to fix the broken copyright system. "Despite this worthy goal," the department wrote to the judge, "the United States has reluctantly concluded that use of the class action mechanism in the manner proposed by the ASA is a bridge too far."

As Downes summarizes what happened during GBS negotiations, it's easy enough to see why DoJ might be concerned:

Somehow, a case about copyright infringement and fair use turned into an agreement to make millions of works available in digital form. While the government "recognizes that the parties to the ASA are seeking to use the class action mechanism to overcome legal and structural challenges to the emergence of a robust and diverse marketplace for digital books," the government's principal objection now is to a more technical question: whether the "broad" scope of the ASA complies with the theory and practice of federal class action law.

That first sentence is key: A narrow issue turned into a sweeping settlement. Calling this a technical issue is certainly correct in some sense, but it's like saying that the mandate of 60Hz 120V AC for U.S. electrical power is technical: True, but that doesn't mean it's trivial.

I think Downes gets it in this paragraph—but I don't believe he wants it:

But a class action is a kind of hammer, and not every complicated legal problem looks enough like a nail to employ it. Here, the parties have not only gone beyond the issues of the original lawsuit, but they have also crafted a settlement that in some sense legislates an orphan works solution that Congress failed to craft. Is that too much innovation for a class action? The Department of Justice "reluctantly" concludes that it is.

Downes' take? Sure, it would be better if legislative issues were acted on by legislation, but...

But in principle, I believe that the elegance of the solution to an otherwise unsolvable problem offered by the ASA makes it a good

candidate for approval. (Elegant, not perfect—but no agreement involving millions of people could ever be perfect.)

In other words: Because it's unlikely that Congress will act, an overbroad settlement should be approved. At the time, I might have agreed. Increasingly, I see that Grimmelmann, Samuelson and others were right: Judge Chin made the only plausible decision.

Monopoly and Antitrust

This handful of items seems primarily related to issues of monopoly, competition and antitrust in GBS, both the original GBS and GBS2. Many other discussions *include* monopoly issues—although GBS seems to be a broader example of what I've seen too often in libraryland, namely a seeming *love* of monopolies as long as they can be perceived as in some way easy, efficient or beneficial.

Google book settlement delayed, DoJ has antitrust concerns

John Timmer posted this at *ars technica* [on April 28, 2009](#). Timmer believed that “Despite the complexity of the settlement, it was on a fast track to approval, with a final thumbs-up scheduled for May [2009].”

Now, it looks like a delay in the decision is inevitable, as opposition to it seems to be rising and the Department of Justice is looking into the antitrust implications of the deal.

The story discusses monopoly issues raised by more than just DoJ. For example, this paragraph begins with objections raised by, among others, Pamela Samuelson's group of academic authors:

So, for example, the agreement as structured could essentially turn Google into the sole rightsholder for orphaned works, which would mean that anyone would have to negotiate with the company over the use of these works. Other objections focus on the fact that Google could control the sale and distribution of out-of-print works, even if the original author decided to release it under a more liberal license. Other [recent objections](#) suggest that the settlement, by giving the search giant control of how the out-of-print works are displayed, could allow the company to censor and selectively display these works, based on community standards or political concerns.

The DoJ's involvement almost seems like an afterthought in this story. It's a good brief roundup of *some* of the objections raised (and Judge Chin's rejection of an Internet Archive attempt to become a party to the suit).

Antitrust and the Google Books Settlement: The Problem of Simultaneity

This article by Eric M. Fraser (University of Chicago Law School and Booth School of Business) was deposited [on June 10, 2009](#) on the Social Science Research Network (SSRN) and was to appear in the September 2010 *Stanford Technology Law Review*. (Hat-tip to Jill Hurst-Wahl for noting the paper on *Digitization 101*.)

It's a 24-page law article; here's the abstract:

Google Books represents the latest attempt at the centuries-old goal to build a universal library. In 2004, Google started scanning books from libraries around the world. Although it made copyright licensing agreements with some publishers, it did not obtain permission from each rightsholder before scanning, indexing, and displaying portions of books from the stacks of libraries. Unsurprisingly, authors and publishers sued for copyright violations. Google settled the class action lawsuit in a sweeping agreement that has raised suspicion from librarians, users, and the government. In this paper, I analyze the antitrust and competition issues in the original and amended settlement agreements. I find that the simultaneous aspects of agreements and pricing pose serious antitrust problems. The settlement effectively gives Google simultaneous agreements with virtually all the rightsholders to in-copyright American books. The original agreement also would have required Google to set prices for books simultaneously. In a competitive market, both agreements and pricing would occur independently. Under current law, however, no potential competitor can make agreements with the rightsholders to orphan works. The simultaneity, therefore, concentrates pricing power, leading to cartel pricing (a problem under § 1 of the Sherman Act) and monopolization (a § 2 problem).

There's little doubt where Fraser stands on the issues. For example, the first sentence of the second paragraph: "The Google Books Settlement Agreement probably violates federal antitrust law." The rest of the article—which I only skimmed and which I may lack the expertise to understand fully in any case—goes into *considerably* more detail on the issues, the flaws and possible alternative courses.

Google's big book case

The slant of this *Economist* story ([September 3, 2009](#), no byline) is clear from the subtitle: "The internet giant's plan to create a vast digital library

should be given a green light.” The first paragraph aims to dichotomize—either you’re a fan or an opponent:

To its opponents, it is a brazen attempt by a crafty monopolist to lock up some of the world’s most valuable intellectual property. To its fans, it is a laudable effort by a publicly minded company to unlock a treasure trove of hidden knowledge. Next month an American court will hold a hearing on an agreement, signed last year by Google and representatives of authors and publishers, to make millions of books in America searchable online. The case has stirred up passions, conflict and conspiracy theories worthy of a literary blockbuster.

Removed from this black-and-white world, many people thought Google Books was a laudable effort that was also monopolistic, and very few opponents failed to credit Google for attempting to “unlock a treasure trove of hidden knowledge.” But that doesn’t make exciting journalism, now does it? (Later in the article, there’s recognition that most critics recognize the potential benefits.)

So how does *The Economist* deal with monopoly issues? It treats the orphan works issue as trivial and claims GBS would *increase* competition. The cartel issue (GBS partners would maintain a legally sanctioned cartel and could raise access prices) is swept away with this argument:

After all, Google has a big economic incentive to ensure that its online library is widely available: it makes most of its money from search advertising, so the more people that use its services, including the online book archive, the better.

Yes and no. Making Google Book *search*—with provisions for buying books—widely available is quite different from keeping institutional *access* affordable. The rest of that paragraph is no more reassuring:

[Google] also has a legal incentive to watch its step. The agreement stipulates that institutional subscription prices must be low enough to ensure that the public has “broad access” to digital books, while at the same time earning market rates for copyright owners. So if lots of libraries refuse to sign up for Google’s service because it is too costly, the company could be slapped with a lawsuit.

None of which negates the antitrust issues, even if the “broad access” term had much enforceable meaning. Realistically, *The Economist*’s argument boils down to a claim that monopoly issues are “theoretical” while the benefits are real.

That is why the court should approve the Google agreement, while at the same time giving stern warning to its signatories that they will be subject to intense regulatory scrutiny for the foreseeable future. If the court rejects the deal, much potentially useful information will remain, quite literally, a closed book.

Monopoly in pursuit of a desirable goal is a good thing. Simple enough. Oh, and Judge Chin can say “regulators will be watching you,” despite the track record of U.S. antitrust in recent decades and the fact that GBS would not *have* an assigned regulator other than the industry-dominated BRR.

Google Books Settlement 2.0: Evaluating Competition

This post, by Fred von Lohmann [on November 19, 2009](#) at *Deeplinks*, is another in EFF’s series of posts analyzing GBS2. He makes the same split of concerns that others have made: the orphan works monopoly and the institutional subscription monopoly, “particularly for higher education.”

Where orphan works are concerned, von Lohmann thinks there’s broad agreement that the monopoly is a bad thing (although *The Economist* waves it aside with a single sentence):

Nobody likes this “only-for-Google” aspect of the settlement—in fact, Google has said that it would support orphan works legislation that would empower the Registry to make the same deal (or even a better deal) with others who want to use these unclaimed works. (Where the claimed books are concerned, in contrast, the Registry will likely ask the rightsholders to appoint it to license companies other than Google. But that still leaves all the unclaimed books out.) The settlement agreement even has a provision that makes it clear that the UWF can license others “to the extent permitted by applicable law”—what amounts to an “insert orphan works legislation here” invitation.

But absent some legislative supplement to the revised Settlement 2.0, it still seems that any other company would have to scan these books, get sued, and hope for a class action settlement. That, of course, is the kind of barrier to entry that any monopolist would envy.

Von Lohmann notes the “worthy question” this raises: If you need legislation to fix the competition problem, shouldn’t the orphan works problem itself be fixed by legislation, not a class action judgment?

Here’s where realpolitik enters the equation. Google correctly points out that Congress has been working on orphan works legislation for years, to no avail. And none of the legislative

proposals came close to the comprehensive solution embodied in the proposed settlement. So the question boils down to a political one: do you believe that approval of Settlement 2.0 will make orphan works legislation more likely, or less likely? Without a crystal ball, it's hard to know.

Here I can't fault EFF or von Lohmann's analysis: It's clear and, I think, fair to all parties involved.

Discussing the Institutional Subscription Database (ISD), the full-access version, von Lohmann assumes that the chief customers are likely to be universities (although I was astonished at the number of assertions at the time that every public library, even the smallest, would be pressured to provide such subscriptions).

The big question is whether, over time, the ISD will become the one database that no university can do without, and the one database with no market substitute (again, because Google will be the only company who can provide a comprehensive corpus without fear of copyright liability, for the reasons explained above). This, of course, is a recipe for monopolistic price gouging, as a group of academic authors led by Prof. Pam Samuelson have pointed out. Over time, universities could face spiraling prices as Google and the Registry conspire to maximize their revenues on the ISD product.

Hmm. Have university libraries faced situations where certain groups of data were felt to be mandatory and without competition, resulting in gouging? I'd think such situations would be, cough, big deals, and libraries would object to having additional pigs at this particular trough. Indeed, the promises in GBS to avoid this situation are less than reassuring if you know much about big deals:

Google and its supporters respond by pointing out that the settlement requires that pricing for the ISD be set with regard to "two objectives: (1) the realization of revenue at market rates for each Book and license on behalf of Rightsholders and (2) the realization of broad access to the Books by the public, including institutions of higher education." The settlement goes on to promise that Google and the BRR "will use the following parameters to determine the price of Institutional Subscriptions: **pricing of similar products and services available from third parties**, the scope of Books available, the quality of the scan and the features offered as part of the Institutional Subscription." [Emphasis added.]

"Similar products"? I wonder what those would be? Digital access to journal articles, possibly? Oh, and GBS didn't give ISD subscribers any

court access to enforce those provisions. “So what we are left with is a ‘trust us’ from Google, the Registry, and their biggest library partners.”

I sometimes give EFF a bad time; I think it goes overboard at times. I read through this piece twice looking for something I could fault. I didn’t find much (but that’s me): I found myself becoming *more* convinced that GBS2 created untenable monopolies as I read it. Here’s the final paragraph of the discussion, after noting DoJ’s investigation and a group of articles arguing the pure legality of the settlement:

But we shouldn’t be satisfied with antitrust law here. This is not just a simple market transaction between commercial entities. Google is building an enormously important public resource, a task it can only undertake with the blessing of a federal court. The public deserves a solution that is not “barely legal,” but that instead encourages real, robust competition. As written, without some modification or legislative adjunct, Settlement 2.0 does not do that.

The Amended Google Books Settlement is Still Exclusive

That’s James Grimmelmann in a relatively brief essay (seven PDF pages) deposited to SSRN [on January 26, 2010](#) and appearing in the *CPI Antitrust Journal* in 2010. The abstract:

This brief essay argues that the proposed settlement in the Google Books case, although formally non-exclusive, would have the practical effect of giving Google an exclusive license to a large number of books. The settlement itself does not create mechanisms for Google’s competitors to obtain licenses to orphan books and competitors are unlikely to be able to obtain similar settlements of their own. Recent amendments to the settlement do not change this conclusion.

Other than noting that the essay isn’t *really* seven pages long (it’s shorter than that), I find that—after reading it and thoroughly enjoying it—I can only say [go read this one](#): It’s too lively and cohesive for me to even attempt to excerpt (I’d have to excerpt almost the whole thing, and what’s the point?).

Why There Can Never Be a Competitor to Google Books

Christopher Mims posted this argument at MIT’s *Technology Review* [on October 18, 2010](#)—a point at which it still seemed plausible that GBS2 would be approved. The subtitle’s clear enough: “Publishers are about to grant Google monopolistic pricing power and permanent exclusivity over countless ‘orphaned’ works.”

To some extent, Mims is excerpting Eric Fraser's article, but he goes further with a paragraph I find offensive and disturbing [emphasis added]:

Here's something Fraser didn't address but I find particularly disturbing: **as more and more libraries disappear, and physical copies of orphaned works become harder to come by**, Google's monopolistic possession of these works will only strengthen. Twenty years from now when e-readers are dirt cheap and we all take digital books for granted, if you find a book on Google Books, who is to say you'll even be able to find a physical copy of it?

Why is it that "more and more libraries" will disappear and physical copies will "become harder to come by"? I guess because of the Digital Inevitability.

Where does Mims come down on all this? He thinks GBS "has implications not just for the future of books, but also for the future of U.S. prosecution of monopolies"—and winds up with this paragraph:

It's also hard to say that Google Books, even if it's a monopoly, isn't a public good in and of itself. The original intention of the indexing project was, after all, to bring all the knowledge hidden in books onto the internet, where it can be searched and integrated into the great hive mind outside of which information is increasingly irrelevant and inaccessible. It's unclear whether or not that will ultimately be good for readers—and not just publishers and the Registry that Google will set up to collect revenue for them.

It's remarkably *easy* to say that GBS2 really doesn't have much to do with the original intention, but never mind. (Oh, and "the great hive mind outside of which information is increasingly irrelevant and inaccessible?" Give me a break.)

Privacy and Confidentiality

While these issues are raised in some items that have already appeared, and should be in some of the library-related items, they're most prominent in this small set, with EFF taking the lead.

Warrants Required: EFF and Google's Big Disagreement about Google Book Search

This Cindy Cohn post [dated August 16, 2009](#) on *Deeplinks* says it right up front:

The central question in the privacy debate that EFF and our partners at the ACLU of Northern California and the Samuelson Law, Technology & Public Policy Clinic at UC Berkeley have been

having with Google about Google Book Search is whether this exciting new digital library/bookstore is going to maintain the strong protections for reader privacy that traditional libraries and bookstores have fought for and largely won.

Libraries and bookstores *have* fought for reader privacy, with ALA and others leading the way, and have indeed largely won that fight. “All we want is for Google to promise to fight for the protections you already have when you walk into a bookstore or a library.”

I’ll quote more, as it’s well stated (and *Deeplinks* has a CC BY license):

One of the most important of those protections is the assurance that your browsing and reading habits are safe from fishing expeditions by the government or lawyers in civil cases. In order to maintain freedom of inquiry and thought, the books we search for, browse, and read should simply be unavailable for use against us in a court of law except in the rarest of circumstances. We have other concerns about Google Book Search as well—concerns and data collection, retention, and reader anonymity—so this won’t end the debate, but safeguards against disclosure are a central point of concern for us.

I’ll pause to bore you with the reminder that these concerns are *not* hypothetical (and that “unless you have something to hide, you shouldn’t care” is a deeply un-American response): *I’ve been there*, as have others in the library field. The FBI *did* conduct fishing expeditions; there is no question about that.

We want Google to promise that it will demand more than a subpoena (which is written by a lawyer and not approved by a judge) or some other legal process that a judge has not approved before turning over your book records. In essence, we asked Google to tell whoever came to them demanding reader information: “Come back with a warrant.”

Honestly, we thought it would be an easy thing for Google to do.

Unfortunately, Google has refused. It is insisting on keeping broad discretion to decide when and where it will actually stand up for user privacy, and saying that we should just trust the company to do so. So, if Bob looks like a good guy, maybe they’ll stand up for him. But if standing up for Alice could make Google look bad, complicate things for the company, or seem ill-advised for some other reason, then Google insists on having the leeway to simply hand over her reading list after a subpoena or some lesser legal process. As Google Book Search grows, the pressure on Google to

compromise readers' privacy will likely grow too, whether from government entities that have to approve mergers or investigate antitrust complaints, or subpoenas from companies where Google has a business relationship, or for some other reason that emerges over time.

We need more than “just trust us” here. EFF has spent the last three years suing AT&T because that company decided, for reasons we still don't know, that it would not stand up for user privacy when the government came knocking. Now, the situations aren't exactly alike—AT&T had a clear legal duty to protect users and demand a warrant, while Google may have more legal options—but that makes it all the more important that Google commit to making the choice to push for a warrant. Reading is deeply personal—as personal as your communications—and we think that Google has a duty to the public to commit to fight for the same level of protection for your bookshelf as for your email.

“Just trust us” is almost never good enough where a publicly owned company is concerned, and especially when that company has a monopoly. If you assume that Internet companies that rely on the good will of citizens to prosper would never undermine privacy, well, back up one letter from G to consider another Internet giant. And consider Eric Schmidt's assertion that privacy is dead anyway—an assertion that, to the best of my knowledge, Google has never disavowed.

By the way, most *good* library systems deal with reader privacy by a means that *assures* that privacy for historical data, warrant or no warrant: They don't retain the history. If it's not there, it can't be subpoenaed. Most ereading systems seem to be going toward the other extreme: Not only is your reading history retained, so are details as to exactly where you currently are in your ebooks.

Google, libraries, and readers' privacy

This post appeared [September 6, 2009](#) on *LibraryLaw Blog*—and it's important to note that it's by Peter Hirtle, not Mary Minow, since the two seemed to be at odds at this point, with Hirtle increasingly sounding like a GBS advocate. (I'm acquainted with Peter Hirtle, and I have considerable respect for him. In this arena, it's possible that his entirely laudable desire to address the orphan works problem was clouding other areas—or, for that matter, that he's right and I'm wrong.)

He notes a revised privacy policy for Google Books and that [EFF didn't find it wholly satisfactory](#). Quoting from the EFF statement (the link in the preceding sentence):

What we asked Google to do was to insist that the most privacy-protective standards be met before disclosing someone's reading history. The position Google has taken instead is that it will follow the few state laws that plainly apply to it already—laws that would bind Google regardless of whether or not Google also wrote about them in its privacy policy. As for the readers living elsewhere, Google says that it will “continue its history of fighting for high standards to protect users,” which is just an aspirational statement, not an enforceable commitment. Google needs to say “come back with a warrant” when law enforcement or civil litigants come knocking for their treasure trove of reader information. This policy does not.

There's a lot more in that post on what Google's policy fails to do—but Hirtle's not impressed:

I would point out that Google's statement is entirely compatible with current library standards for confidentiality in licensed resources.

Which may be true—although it's not quite. The model license Hirtle quotes includes this statement: “Raw usage data, including but not limited to information relating to the identity of specific users and/or uses, shall not be provided to any third party.” It doesn't say “except where legal processes are followed.” It says “shall not.” That's a *huge* difference, especially since “legal processes” includes subpoenas, which do not involve a judge's assent. (Another e-resource license is much inferior in this regard.)

Hirtle omits books from his discussion—and books are where libraries have the strongest privacy protections. His final paragraph:

The bottom line: Google is more than compliant with current library standards for 3rd-party privacy protection. EFF argues that “Given the important free expression interests at stake and the long history of protecting reader privacy by libraries and bookstores, readers need a durable guarantee of protection enforceable by a court.” No library has been demanding such a guarantee before now. One has to wonder if the current criticism of Google wouldn't be better directed at libraries and their privacy requirements when working with outside vendors.

Libraries *achieved* that level of protection for books. The first license quoted also fully protects reader privacy by forbidding *any* distribution of non-aggregated data to third parties. On the other hand, Hirtle is certainly correct in saying that libraries should hold e-resource vendors to the same standards they use for book data (and, by the way, they

should consistently uphold their own standards). In practice, the loophole that a reader logged in to their Google account might have personal data logged is a loophole big enough to drive several FBI squadrons through, as Google pushes more and more services to assure that you're always logged in, whether you're aware of it or not. (Do *you* explicitly log out of Gmail after each session? Really?)

The comment stream is worth reading and consists mostly of an ongoing disagreement between Mary Minow and Peter Hirtle.

Google Books Settlement 2.0: Evaluating Privacy

Here we are again with Fred von Lohmann on *Deeplinks*, this time [on November 23, 2009](#). He notes the level of information that Google *might* have under GBS2 and what it means for readers:

The products and services envisioned by the proposed settlement will give Google not only an unprecedented ability to track our reading habits, but to do so at an unprecedented level of granularity. Because the books will be accessed on Google's servers, Google will not only know what books readers search for and access, but will also know which pages they read, how long they stayed on each page, what book they read before, and which books they access next. This is a level of reader surveillance that no library or bookstore has ever had.

Readers who feel surveilled will be chilled in their freedom of inquiry. As Supreme Court Justice William O. Douglas observed in 1953, "Once the government can demand of a publisher the names of the purchasers of his publications . . . [f]ear of criticism goes with every person into the bookstall . . . [and] inquiry will be discouraged." Or as Author Michael Chabon put it: "If there is no privacy of thought—which includes implicitly the right to read what one wants, without the approval, consent or knowledge of others—then there is no privacy, period."

There are other intrusions—and here EFF notes some numbers: at least 200 attempts by law enforcement to get patron reading information just between 2000 and 2005. Von Lohmann provides a laundry list of privacy failures in GBS2; it's a fairly impressive list. He concludes:

For all of these reasons, in its present form and without further affirmative steps by Google either in the context of the settlement or outside it, the proposed Settlement 2.0 makes Google Books a threat to reader privacy, which in turn is a serious a down-side that must be weighed against the settlement's potential benefits.

GBS: Jones and Janes on Anonymity in a World of Digital Books

Here's a sad case where I can link to James Grimmelmann's [December 22, 2010 post](#) at *The Laboratorium*, which touts and links to "[Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read](#)" by Elizabeth Jones and Joseph Janes—but I can't discuss the paper itself because these two iSchool faculty chose to publish in a toll access journal that only provides "free" guest access under certain conditions, among them institutional affiliation. Those without affiliation (like me) apparently have no reason to read the article. (Since I could probably get the article upon registration by offering an institutional affiliation that I don't have...well, I'm not going to do that.)

Grimmelmann says "It is the most careful and sustained analysis to date of the privacy issues surrounding the proposed settlement" but about all I can do is quote the abstract, as he does:

With its Books project, Google has made an unprecedented effort to aggregate a comprehensive public-access collection of the world's books. If successful, Google's collection would become the world's largest and most broadly accessible public book collection—indeed, project leaders have frequently spoken of their desire to create a "universal library" (Toobin 2007). Still, the Google "library" would differ from established contexts for the provision of free, public access to reading materials—like public libraries—along several policy-related dimensions, of which perhaps the most glaring is its treatment of reader privacy. This paper teases out the specific differences in reader privacy protections between the American public library and Google Books, and what those differences might mean for the values and goals that such contexts have historically embodied. Our analysis is structured by Helen Nissenbaum's "contextual integrity decision heuristic" (2009), which focuses on revealing changes in informational norms and transmission principles between prevailing and novel settings and practices. Based on this analysis, we recommend a two-pronged approach to alleviating the threats to reader privacy posed by Google Books: both data policy modifications within Google itself and inscription of privacy protections for online reading into federal or international law.

Many of my readers may have institutional access to this article or be ethically comfortable in filling in institutional information, in which case this might be a great article (although, frankly, in December 2010 proposing modifications to GBS was almost certainly a nonstarter). I couldn't say.

Google Books Decision: “The Privacy Concerns are Real”

This relatively brief March 22, 2011 post by Cindy Cohn at *Deeplinks* notes (correctly) that Judge Chin’s decision striking down GBS *did* mention privacy concerns—and also that he did not find these concerns to be sufficient to reject the proposed settlement. Two key paragraphs:

While noting that “[T]he privacy concerns are real,” the court decided that they were not a basis, in themselves, to reject the proposed settlement. It noted that the settlement contained privacy protections for Rightsholders and also noted that Google had “committed” to certain safeguards for readers, while acknowledging that those were voluntary only. The court closed with a strong nudge to Google: “I would think that certain additional privacy protections could be incorporated, while still accommodating Google’s marketing efforts.”

We look forward to continuing our discussions with Google about implementing additional privacy protections in whatever form the Google Books project takes as it moves forward. In the meantime, EFF and the ACLU are also working together on digital book privacy legislation in California, which should be introduced shortly. The proposed law, which partially grew out of our negotiations with Google, will extend to digital booksellers and libraries the longstanding privacy protections against overreaching government and civil litigation demands for information about readers.

That may be as good a place as any to close this section, noting that privacy and confidentiality show up later in this roundup.

The Public Domain, Open Access, Copyright and Fair Use

While the latter two topics here are at the heart of nearly all of GBS and commentaries on it, I have a few items specifically focusing on these topics, so I’m lumping them together here in that order.

The Google Book Settlement and the Public Domain

We’re back to *LibraryLaw Blog*, a fairly long Peter Hirtle post [on April 9, 2009](#). It’s an expansion on a quick response he gave when “a colleague wrote to ask what I thought of [GBS] procedures for identifying public domain books.”

My quick assessment: the settlement specifies procedures that are likely to identify most public domain works published in the

United States. It is less helpful for foreign publications that may have entered the public domain; they are largely absent from the process. Unfortunately, because this is part of litigation rather than legislation, no one else can take advantage of the results of the process—it moves us no closer to having a growing public domain. What is unknown is to what extent Google will want to remove titles from the licensed products and make them freely available to the public.

That's the core; the rest of the post is details. Peter Hirtle is an expert in this area and the discussion is eminently worth reading—even with the failure of GBS itself. I don't have much to add, so I'll just suggest that those interested in the public domain and some of the issues involved in trying to identify what's part of it will find Hirtle's post worthwhile.

Another idea for building OA into the Google Book Settlement

If that title (from [a June 17, 2009 post](#) by Peter Suber at *Open Access News*) seems to demand a referent, it's there in one of Suber's bullet points regarding the post he's linking to, "Google Book Search Settlement: Foster Competition, Escrow the Scans" (by Peter Eckersley [on June 11, 2009](#) at *Deeplinks*). He notes two other proposals that would build OA support into GBS.

I would comment here, but I struggled with identifying anything in Eckersley's post that deals directly with OA. That probably means my understanding of OA is lacking when compared to Suber, which seems likely. You may spot the connections that I'm missing.

Revised Google Book settlement: what it means for OA

Suber revisits GBS, this time GBS2, in this [November 16, 2009](#) *Open Access News* post. He notes the most directly-related changes: That the Book Rights Registry would "facilitate Rightsholders' wishes to allow their works to be made available through alternative licenses for Consumer Purchase, including through a Creative Commons license" and that it's now clear that Rightsholders would be free to set the consumer purchase price of their books at zero.

Suber also notes that GBS2 does *not* include these provisions for the Unclaimed Works Fiduciary, the body that would actually have dealt with true orphan works. That is, it would *not* have had the ability to make true orphan works open access by setting the price at zero (or reduce copyright restrictions by using a CC license).

Open access and the Google book settlement

That's the lead article in the [December 2, 2009](#) *SPARC Open Access Newsletter*—and again it's by Peter Suber. (If it seems odd that every item

on OA comes from Peter Suber: It shouldn't. Especially where something beyond current refereed science, technology and medicine journal articles are concerned.)

Suber notes that many other people were looking carefully at GBS2 and that there are large questions in several areas—all of which he's ignoring because OA is his specialty, not because they're not important. Suber's key points:

(1) The first point to make is that OA was never an issue in the lawsuit. Google wasn't scanning copyrighted books in order to make them OA, and the plaintiff groups didn't sue Google because they thought it was making them OA or planning to make them OA.

However, Google's wide-ranging book-scanning program did overlap with OA. For example, Google was scanning public-domain books and making them at least gratis OA. But the lawsuit raised no objection to the public-domain scans or their terms of access. When the lawsuit was filed, Google suspended its scanning of copyrighted books, but continued its scanning and posting of public-domain books without objection from any quarter...

Bottom line: if Google had never been sued, or if it had won the suit outright, without having to settle, we still wouldn't have OA to the scanned, copyrighted books which are the subject of the suit. In that sense, the lawsuit did not prevent OA to any class of books and the settlement is not a retreat from an earlier plan to provide OA.

(2) If there's an exception, it's an attenuated sort. Both the original and amended settlement provide for free online access from a small number of terminals in libraries (Sections 1.117 and 4.8.a.i) to at least 85% (Section 7.2.e.i.1-2) of the corpus of otherwise non-OA digital books...

We shouldn't call this OA, however. These provisions don't make any books OA. They merely give users a kind of special access to non-OA books.

This exception has the approval of the plaintiffs, of course, or it would not appear in the settlement. We could say that it's analogous to the accommodation authors and publishers have made to the existence of free lending libraries. But before we get too comfortable with that analogy, we should remember that the Authors Guild has not fully accommodated the existence of free lending libraries. As recently as 1987 it demanded "a government-funded royalty paid to authors of books borrowed from libraries."

An aside to certain librarians celebrating their eligibility to join Authors Guild: Is this really a group you want to support? I know my answer...

Moreover, far more citizens have free access to print books through free lending libraries than will have free online access to digital books through the small number of privileged library terminals....

He reminds us of the limits on GBS' largesse: One terminal for every 10,000 FTE students at universities, one for every 4,000 at community colleges, one *per building* for public libraries (which comes down to an average of one for roughly every 18,000 citizens).

(3) Another attenuated sort of exception is that both versions of the settlement by default allow Google to display up to 20% of any copyrighted book it scans under the program (Section 4.3.b.i.1). This is a larger portion than the tiny snippets Google displays today. When OA people say that a text is OA, they mean that the full-text is OA. In that sense, it would be misleading to call the 20% slices "OA texts". But no matter what terms we use to describe them, these slices are gratis OA and larger than the snippets that came before.

There's a *lot* more here—I've just scratched the surface.

GBS: Samuelson on the Settlement as Copyright Reform

In his role as the Peter Suber of GBS (a comparison that's probably unfair to both of them), James Grimmelman posted this item on *The Laboratorium* [on September 30, 2010](#). He's pointing to Pamela Samuelson's "[The Google Book Settlement as Copyright Reform](#)" and says Samuelson's long, distinguished history of engagement with copyright reform efforts "gives this paper an unusually synoptic view of the copyright issues raised by the lawsuit and settlement." He calls the paper "a gold standard of sophisticated analysis." The abstract:

An intriguing way to view the proposed settlement of the copyright litigation over the Google Book Search (GBS) Project is as a mechanism through which to achieve copyright reform that Congress has not yet and may never be willing to do. The settlement would, in effect, give Google a compulsory license to commercialize millions of out-of-print books, including those that are "orphans" (that is, books whose rights holders cannot readily be located), establish a revenue-sharing arrangement as to these books, authorize the creation of an institutional subscription database that would be licensed to libraries and other entities, resolve disputes between authors and publishers over who owns

copyrights in electronic versions of their books, provide a safe harbor for Google for any mistakes it might make in good faith as to whether books are in the public domain or in-copyright, and immunize libraries from secondary liability for providing books to Google for GBS, among other things.

This Article explains why certain features of U.S. law, particularly copyright law, may have contributed to Google's willingness to undertake the GBS project in the first place and later to its motivation to settle the Authors Guild lawsuit. It then demonstrates that the proposed settlement would indeed achieve a measure of copyright reform that Congress would find difficult to accomplish. Some of this reform may be in the public interest. It also considers whether the quasi-legislative nature of the GBS settlement is merely an interesting side effect of the agreement or an additional reason in favor or against approval of this settlement.

If Grimmelmann says (as he does) "highly recommended," who am I to say otherwise? I will note that it's an 84-page article—and that it was revised in April 2011, so that the third major section is headed "Should the GBS Settlement Have Been Approved?"—with *Have Been* rather than *Be*. I have *not* read Samuelson's article in full (although I've downloaded it); I'm passing along Grimmelmann's recommendation and noting that this sort of analysis is most definitely still worthwhile even after GBS was rejected.

Professional Readings: Sag's The Google Book Settlement and the Fair Use Counterfactual

This brief post by Joe Hodnicki [on August 19, 2009](#) at *Law Librarian Blog* simply points to and offers the abstract of Matthew Sag's [article with that title](#). I have *not* downloaded that 47-page article (available from SSRN)—and, as with the next two pieces, you might appropriately think of this as a brief extension of the long, long fair use roundup in *C&I* June 2012 and July 2012. The abstract (emphasis added):

This Article compares the Google Book Search Settlement to the most likely outcome of the litigation the settlement resolves. It argues that Google was never likely to receive the courts unqualified approval for its massive digitization effort and that **the most likely outcome of the litigation was that book digitization would qualify as a fair use subject to an opt-out**. Accordingly, the aspects of the proposed settlement which allow Google to continue to operate its book search engine in its current form should not be controversial; they essentially mirror the court's

most likely fair use ruling if the case had gone to trial. In effect, the opt-out that fair use would likely have required has been replaced by the ability of copyright owners to opt out of the class-action settlement.

In the wake of the proposed Settlement, the Google Book debate has shifted away from the merits of book digitization, and refocused on questions of commoditization and control. This Article highlights four critical areas in which the Settlement differs sharply from the predicted fair use ruling. First, the Settlement permits Google to engage in a significant range of uses including the complete electronic distribution of books that go well beyond fair use. Second, the Settlement provides for initial cash payments by Google to the copyright owners and a fairly generous revenue sharing agreement, neither of which would have been required under a fair use ruling. Third, the agreement creates a new set of institutional arrangements that will govern the relationship between Google and the copyright owners covered by the Settlement. The foundations of this new institutional framework are the Settlement agreement itself, the creation of a collective rights management organization called the “Book Rights Registry” and the “Author Publisher Procedures”. The fourth area in which the Settlement differs from the likely fair use outcome relates to the accessibility, commoditization and control of orphan works.

What the Google Books Decision Said About Fair Use

We jump forward to post-decision discussions, including this *ARL Policy Note* posted by Brandon Butler [some time around April 9, 2011](#) (that’s when I tagged it in Diigo; the time stamp on the piece is “1 year ago”).

As pundits and participants weigh in on the meaning of Judge Chin’s rejection of the Google Books settlement, it is important that one thing remain crystal clear: Judge Chin did not rule on the issue at the heart of the original dispute, whether it was a fair use to scan in-copyright books to facilitate search and to display snippets from those books in search results. That question remains wide open.

Indeed it does: The rejection was based on fairness as a class action settlement.

While there was neither a holding nor even a real discussion of the original fair use issue, Judge Chin’s opinion did include a few conflicting asides (or *obiter dicta* in lawyer-speak) on the issue. On

page 25, Judge Chin characterized the original project as involving “an indexing and searching tool,” a characterization that, if anything, favors the argument that Google’s activities were fair use. After all, a similar “indexing and searching tool,” Google’s Internet search engine, is fairly well established as a fair use despite its unauthorized copying of entire Internet websites as part of the indexing process. And creating a search tool is a transformative use that will not supersede the original works that are copied, a powerful argument for fair use. But later, on page 27, Judge Chin described Google’s activities as “blatant, wholesale copying,” then quotes objectors characterizing Google’s book scanning as a “shortcut” in “disregard of authors’ rights.” Perhaps Judge Chin was just channeling the objectors here, rather than expressing his own views, but in any case, these tossed-off and inconsistent characterizations do not constitute a legal holding.

So, as we all work to decide what this latest twist in the Google Books saga means for our communities, we should keep one thing in mind: Google’s original fair use argument for scanning and snippet display remains persuasive, and has yet to be tested in court.

An important point. While Google could have considerably expanded the general understanding (and likely use of) fair use by successfully defending itself in court—and may still do so—the rejection of GBS2 had *nothing* to do with fair use. And, indeed...

Google Should Stand up for Fair Use in Books Fight

So says Timothy B. Lee in [this March 22, 2011 post](#) at *Freedom to Tinker*. Lee argued early on that Google’s scanning and snippet displays were legitimately fair use and still thinks that’s right. His summary of a three-year interruption is pretty good for one paragraph:

Unfortunately, in 2008 Google saw an opportunity to make a separate truce with the publishing industry that placed Google at the center of the book business and left everyone else out in the cold. Because of the peculiarities of class action law, the settlement would have given Google the legal right to use hundreds of thousands of “orphan” works without actually getting permission from their copyright holders. Competitors who wanted the same deal would have had no realistic way of doing so. Googlers are a smart bunch, and so they took what was obviously a good deal for them even though it was bad for fair use and online innovation.

Well, it wasn't *just* [a segment of] the publishing industry, it was also [a tiny segment of] authors. But it's not a bad way to look at it.

Lee thinks the *rejection* of GBS might strengthen Google's fair use argument.

Fair use exists as a kind of safety valve for the copyright system, to ensure that it does not damage free speech, innovation, and other values. Although formally speaking judges are supposed to run through the famous four factor test to determine what counts as a fair use, in practice an important factor is whether the judge perceives the defendant as having acted in good faith. Google has now spent three years looking for a way to build its Book Search project using something other than fair use, and come up empty. This underscores the stakes of the fair use fight: if Judge Chen ruled against Google's fair use argument, it would mean that it was effectively impossible to build a book search engine as comprehensive as the one Google has built. That outcome doesn't seem consistent with the constitution's command that copyright promote the progress of science and the useful arts.

The first comment is from James Grimmelman, who offers a quick prediction:

My prediction is that the settlement will be opt-out as to scanning and searching, and opt-in as to full-text, thus protecting Google from copycat lawsuits (which would need to seize control of this case from its current counsel).

Libraries and Metadata

We now come to the largest section and the one closest to my heart. I will admit that I anticipated seeing a *lot* more "libraries can get rid of all those boring books and rely on Google" sentiments from both academic and public librarians, since I seem to remember being galled by dozens of such absurdities at the time. Either I managed to avoid tagging them (quite possible) or my memory's faulty (even more possible), but I don't see a lot of that here. What's here is a combination of reporting and commentary, most of it from 2009 and 2010 with a few items from 2011.

Library Privileges (Fees May Apply)

How better to start than with the estimable Barbara Fister, writing here in "Peer to Peer Review" [posted April 23, 2009](#) at *Library Journal*. Fister's noting the stream of reactions to the original GBS—some of them covered in [my March 2009 GBS roundup](#)—and was struck by something about many of the reactions:

[O]ne thing that has struck me isn't about Google at all. It's the undercurrent of frustration and hostility toward higher education, academic libraries, and private research libraries like Harvard's in particular. What right do libraries have, the critics ask, to disapprove of a project that liberates books from the gloomy stacks and provides access to the people, not just the elite who can afford obscenely high tuition and run the gauntlet of highly selective admission?

She cites a prime example (a post that's wrong on several counts, even if the original GBS had been approved), notes that Google isn't *really* free and some of the virtues of public libraries and adds:

Still, it's obvious that we haven't made this clear to a lot of the citizenry, in words or in actions. And some of the invective I've come across in comments at blogs and newspapers is startlingly vitriolic, a populist backlash against academia's claims.

Fister notes the reality of most academic libraries, especially private institutions. I can't walk into Harvard's libraries and borrow a book, for example—indeed, I'm not sure I could even go look at a book on site without registering and possibly paying.

Er, but...we're all about the public good, right? At least that's the claim we make when we criticize a private corporation for monetizing library collections. We're superior because we're not in it for the money. Our materials are there to support research—just not yours. We're here for our immediate community, and you're not part of it. Go home to your local library (if you have one—and lots of U.S. citizens live in places that don't have them) and request our stuff by interlibrary loan (if it's available; fees may apply).

After noting exceptions (I'm aware that I have access to a number of quality academic library collections through Link+, a Northern California union catalog of sorts, but I'm not sure that's at all typical), Fister says:

Still, you can't get around the fact that for many people, academic libraries are perceived as a luxury accessory for those who can afford to go to college. Shaking our fingers at those who don't recognize the dangers of Google's commodification of culture (something I confess I do regularly) had better take into account the ways our library practices are a product of a similar commodification of American higher education. If we seriously think we serve the common good, we need to examine exactly how we do that in practical ways.

Library Associations File Amicus Brief for Google Book Search Settlement

That's Peter Murray, the *Disruptive Library Technology Jester*, reporting [on May 4, 2009](#), including a link to the [22-page brief](#) filed on behalf of ALA, ACRL and ARL. If you want to read the brief, don't be put off by the number of pages—that's 22 double-spaced pages, not really all that much text, and with Jonathan Band as lead attorney, it's readable prose. The brief did not oppose GBS but did raise a number of library-related issues. From the brief:

The Settlement, therefore, will likely have a significant and lasting impact on libraries and the public, including authors and publishers. But in the absence of competition for the services enabled by the Settlement, this impact may not be entirely positive. The Settlement could compromise fundamental library values such as equity of access to information, patron privacy, and intellectual freedom. In order to mitigate the possible negative effects the Settlement may have on libraries and the public at large, the Library Associations request that this Court vigorously exercise its jurisdiction over the interpretation and implementation of the Settlement.

Specific issues raised and discussed:

- The likelihood that the institutional subscription database (ISD) would be seen as indispensable, but also a monopoly open to abuse.
- The possibility that Google could at some point seek a “profit maximizing price structure” for the ISD “that has the effect of reducing access”—especially since the model for the pricing, online journal packages, has distinctly had fees that grow so high only the wealthiest libraries can subscribe.
- High ISD pricing could heighten inequalities among libraries—including the bizarre situation where K12 school libraries might be able to afford the ISD but college libraries might not.
- GBS does not protect patron privacy and contains provisions that appear to undermine privacy.
- GBS could potentially limit intellectual freedom.
- GBS could “frustrate the development of innovative services.”
- These shortcomings can be dealt with through rigorous oversight of GBS implementation. (The brief includes half a dozen specific elements of such oversight.)

Murray's post does a fine job of excerpting key elements of the brief, and even at this late date it's worth seeing where library groups were coming from.

Two on Michigan's Changes

Peter Murray offers another *Disruptive Library Technology* post [on May 22, 2009](#), this one entitled “Interesting Bits in the Univ of Michigan Amendment to Google Book Search Agreement,” and Jonathan Band issues one in his “A Guide for the Perplexed” series, “Part II: The Amended Google-Michigan Agreement,” which seems to have appeared [on June 12, 2009](#) (based on PDF properties). The [amended agreement](#) between the University of Michigan (one of the primary Google Library Project partners) and Google addresses issues raised in the lawsuits and would govern the relationship between the two parties if GBS was approved.

Murray notes that “there are definitely more lawyers involved now”—such that while the original Michigan-Google contract was “basically 10 pages long,” the amendment is 36 pages and incorporates by reference the entire GBS.

The amendment also liberally copies and pastes entire 200-word clauses in sections and includes sentences that are over 400 words long. If these lawyers are paid by the hour, they made out like bandits—as have the pain reliever companies from selling all of the medicines to cope with the headaches that come from reading it.

I'll try not to cause you the same pain as I point out the good bits.

Some of the good bits may be moot with the failure of GBS, but it's worth noting what Murray found worth mentioning. There's an explicit process for modifying the library's digital copy of scanned books that are identified as being in the public domain. The revised agreement appears to include Michigan's Special Collections, explicitly excluded from the original. There's a provision for challenging ISD pricing—but also promises to donate significant sums to the National Federation for the Blind if prices are *not* challenged. (In any case, Michigan would have 25 years of free access as part of GBS.) There's also a promise to provide at least \$5 million in aggregate to support centers for research use of the body of scanned material.

Band's “Guide for the Perplexed” is brief (seven double-spaced pages) and focuses on changes that apply to all of Google's partner libraries, not just Michigan. Those include mechanisms for pricing review (by a third party) and arbitration, information to be provided by Google to the participating libraries for the works scanned (including their apparent copyright status and whether they're being displayed) and a few other things. Also an interesting read.

Librarians Fighting Google's Book Deal

Whether under this title (the actual title on the story) or under the webpage title “Librarians vs. Google: Fighting the Web Giant's Book

Deal,” this [June 17, 2009 story](#) by Janet Morrissey at *Time* clearly paints libraries as *enemies* of GBS.

Critics of Google’s book-searching agreement with publishers and authors were cheered last week when antitrust regulators in the Justice Department set their sights on the search giant’s publishing deal, demanding more information.

“This is a monumental settlement that’s at stake, and for the government to show this kind of attention is heartening,” says Lee Van Orsdel, dean of university libraries at Grand Valley State University. “The increased scrutiny on the part of the DOJ tells us that our concerns are resonating far beyond the library community,” concurs Corey Williams, associate director in the office of government relations at the American Library Association.

Those are the first two paragraphs. Oddly enough, I do not read the second paragraph as librarians *fighting* Google. I see it as librarians asking for appropriate concerns to be addressed. But the next paragraph reaffirms *Time*’s view that it’s a Battle Royale:

Goliath Google facing off against a legion of librarians and, possibly, the U.S. Justice Department—now there’s a fight.

Morrissey then says GBS “once appeared to be a sure bet for rubber-stamp approval” but thanks to that “angry opposition” its future might now be in question. As she goes on to oversimplify GBS itself, there’s the apparently mandatory reference to “dusty shelves in university libraries” that GBS rescues books from. After some more description, the article does include a paragraph that should clarify the reason for *one* set of library concerns:

The library community recalls with horror the pricing fiasco that occurred when industry consolidation left academic journals largely in the hands of five publishing companies. The firms hiked subscription prices 227% over a 14-year period, between 1986 and 2002, forcing cash-strapped libraries to drop many subscriptions, according to Van Orsdel. “The chance of the price being driven up in a similar way (in the Google deal) is really very real,” she says.

As a humanities person, I’ll note that libraries also substantially reduced budgets for monographs in order to keep feeding the insatiable Elsevier & Friends.

The article does note that it’s not just librarians—that “academics, consumer advocates and even a few authors” were opposed to GBS as it stood. (“A few authors” is a nice way of belittling all authors who were opposed, including 6,000 who opted out and the academic authors involved in Pamela Samuelson’s brief.) And of course she quotes

“supporters”—such as, ahem, the executive director of the Authors Guild and the engineering director of Google. Oh, and such other disinterested parties as the law firm that’s representing the Authors Guild and the CEO of one of the publishers involved. Indeed, the only supporter who wasn’t directly involved in the case was Paul Courant—but you can’t really call the University of Michigan not directly involved, can you?

Without the combative headline, this would be a better story. Still, it’s interesting that the *only* supporters Morrissey managed to quote were all directly or indirectly involved in the settlement.

The undisclosed danger to libraries in the Google Books Settlement

That’s Peter Hirtle, writing on [August 16, 2009](#) at *LibraryLaw Blog*. He focuses on an aspect of GBS he hadn’t seen discussed much:

I have been surprised at the lack of discussion in the library community about what I feel is one of the most problematic features of the settlement: printing fees in the Public Access Service. The Public Access Service is the free license that every public library can receive that allows that library to access the proposed books database from one of the library’s computers. Users are allowed to view the entire text of the book (unlike the Consumer Purchase model, which only allows you to see up to 20% of the book without paying), but they are not allowed to download the book. Users can, however, print out pages from the book.

Here is the kicker: if the library charges a fee for printing (and how many libraries can allow users to print for free?), then they are required by Section 4.8(a)(ii) of the Agreement to charge users for the printing. Google will collect the money on behalf of libraries and pass it on to the Registry. Google has agreed to pay the cost of the printing for the first five years or \$3 million, whichever comes first.

It is standard practice in many libraries to charge for the cost of paper and toner associated with printing from networked resources. I cannot think of a single licensed resource, however, that also wants libraries to pay a use fee for that printing. It is the equivalent of not only having users pay for costs of photocopying, but also having to send a royalty check to the Copyright Clearance Center for every page they print. And note that there is no provision for fair use in this requirement—printing even one page will result in the payment of a royalty to the Books Rights Registry.

Hirtle also notes that this provision alone could raise enormous privacy issues and that the provision appears to “overturn almost 75 years of law

and practice” that have protected libraries from being required to collect royalties when patrons copy materials.

Privacy, anti-trust, and orphan works are important issues. But am I wrong in thinking that this innocuous-sounding little clause in the middle of the Agreement may do more to change the way libraries operate than other element of the Settlement?

It’s a good question even if it has been rendered moot. You may find the comments interesting.

Press Round-Up: UC Berkeley Conference Regarding Google Book Search Settlement

Because this was a conference (or symposium) and, as a result, you’re mostly dealing with second-hand reporting, I don’t plan to offer detailed notes on the various links from this [August 28, 2009 ResourceShelf](#) report. You may find it interesting as a set of comments on a particular event, one that included some high-profile doubters and supporters.

Marcus Banks attended the symposium and reported on it [on August 30, 2009](#) at *Marcus’ World*, but most of the post consists of Banks’ quick summary of GBS elements and his feelings about them.

You can read much more in depth than I have, but ultimately the debate boils down to this question: does the settlement facilitate a monopolistic cartel between Google and major publishers, or does it open up access to the world’s literature in a truly innovative and unprecedented way?

The answer is: both. So where you stand on the settlement depends on how you weigh the relative risks and benefits.

In the end, Banks comes down on the side of approval “even though there is a real risk that Google could act monopolistically.” He thinks the Library of Congress *should* have been doing this project—but it didn’t. Therefore,

[A]s with so many other things in life and in libraries, with Google Books we’ve reached the point where we shouldn’t let the perfect be the enemy of the good.

I can’t fault Banks for that; it’s how I felt at the time.

Google Books: A Metadata Train Wreck

This post [on August 29, 2009](#) by Geoff Nunberg at *Language Log* also relates to the Berkeley symposium (he calls it a conference) and specifically Nunberg’s evaluation of the metadata for GBS scans.

My presentation focussed on GB’s metadata—a feature absolutely necessary to doing most serious scholarly work with the corpus. It’s

well and good to use the corpus just for finding information on a topic—entering some key words and barrelling in sideways. (That’s what “googling” means, isn’t it?) But for scholars looking for a particular edition of *Leaves of Grass*, say, it doesn’t do a lot of good just to enter “I contain multitudes” in the search box and hope for the best. Ditto for someone who wants to look at early-19th century French editions of *Le Contrat Social*, or to linguists, historians or literary scholars trying to trace the development of words or constructions: Can we observe the way happiness replaced felicity in the seventeenth century, as Keith Thomas suggests? When did “the United States are” start to lose ground to “the United States is”? How did the use of propaganda rise and fall by decade over the course of the twentieth century? And so on for all the questions that have made Google Books such an exciting prospect for all of us wordinistas and wordastri. But to answer those questions you need good metadata. And Google’s are a train wreck: a mish-mash wrapped in a muddle wrapped in a mess.

The rest of the fairly long post elaborates on that point—loads of books misdated as 1899, a book on Peter Drucker dated 1905 and more. He says these errors are endemic: At the time, a search on “internet” in books written before 1950 resulted in 527 hits.

Or try searching on the names of writers or famous restricting your search to works published before the years of their birth. You turn up 182 hits for Charles Dickens, more than 80 percent of them misdated books referring to the writer as opposed to someone else of the same name. The same search turns up 81 hits for Rudyard Kipling, 115 for Greta Garbo, and 29 for Barack Obama. (Or maybe that was another Barack Obama.)

Nunberg says Google’s Dan Clancy “said that the erroneous dates were all supplied by the libraries,” which Nunberg doesn’t believe, especially given the errors for books correctly dated in the libraries’ catalogs.

Most of the misdatings are pretty obviously the result of an effort to automate the extraction of pub dates from the OCR’d text. For example the 1604 date from a 1901 auction catalogue is drawn from a bookmark reproduced in the early pages, and the 1574 dating (as of this writing) on a 1901 book about English bookplates from the Harvard Library collections is clearly taken from the frontispiece, which displays an armorial bookplate dated 1574:

It’s not just dates. He notes classification errors—e.g., *Moby-Dick* listed under “Computers.” Google blames the libraries for this as well—which

really doesn't work because the Google subject classifications are BISAC categories, not LC subject headings.

But whether it gets the BISAC categories right or wrong, the question is why Google decided to use those headings in the first place. (Clancy denies that they were asked to do so by the publishers, though this might have to do with their own ambitions to compete with Amazon.) The BISAC scheme is well suited to organizing the shelves of a modern 35,000 foot chain bookstore or a small public library where ordinary consumers or patrons are browsing for books on the shelves. But it's not particularly helpful if you're flying blind in a library with several million titles, including scholarly works, foreign works, and vast quantities of books from earlier periods. For example, the BISAC "Juvenile Nonfiction" subject heading has almost 300 subheadings, including separate categories for books about "New Baby," "Skateboarding," and "Deer, Moose, and Caribou." By contrast, the "Poetry" subject heading has just 20 subdivisions in all. That means that Bambi and Bullwinkle get a full shelf to themselves, while Schiller, Leopardi, and Verlaine have to scrunch together in the lone subheading reserved for "Poetry/Continental European." In short, Google has taken the great research collections of the English-speaking world and returned them in the form of a suburban mall bookstore.

There's still more in a post rife with screenshots, including the rather lovely *The Mosaic navigator: The essential guide to the interface*, by Sigmund Freud and Katherine Jones, published in 1939. I wonder what Freud thought of the web browser?

Nunberg says Google's aware of the errors (even as it seems intent on blaming its library partners) and plans to fix them—but "they've acknowledged that this isn't a priority." He doesn't believe one-at-a-time corrections as errors are reported can work in such a large corpus with so many errors.

Some 80 comments, well worth reading (I'll admit to not reading them all—quite a few are long), including a Google metadata person's admission that the errors don't number in the hundreds of thousands: there are *millions* of errors.

Google Book Search: A Disaster for Scholars

You could think of this [August 31, 2009 piece](#) by Geoff Nunberg at *The Chronicle of Higher Education* as being a more formal version of the post above. Nunberg offers enough context to make it a useful article for *CHÉ's* wider readership. Nunberg concludes with two optimistic

paragraphs—and it's fair to remind readers that, in August 2009, most people assumed GBS *would* be approved:

I'm actually more optimistic than some of my colleagues who have criticized the settlement. Not that I'm counting on selfless public-spiritedness to motivate Google to invest the time and resources in getting this right. But I have the sense that a lot of the initial problems are due to Google's slightly clueless fumbling as it tried master a domain that turned out to be a lot more complex than the company first realized. It's clear that Google designed the system without giving much thought to the need for reliable metadata. In fact, Google's great achievement as a Web search engine was to demonstrate how easy it could be to locate useful information without attending to metadata or resorting to Yahoo-like schemes of classification. But books aren't simply vehicles for communicating information, and managing a vast library collection requires different skills, approaches, and data than those that enabled Google to dominate Web searching.

That makes for a steep learning curve, all the more so because of Google's haste to complete the project so that potential competitors would be confronted with a *fait accompli*. But whether or not the needs of scholars are a priority, the company doesn't want Google's book search to become a running scholarly joke. And it may be responsive to pressure from its university library partners—who weren't particularly attentive to questions of quality when they signed on with Google—particularly if they are urged (or if necessary, prodded) to make noise about shoddy metadata by the scholars whose interests they represent. If recent history teaches us anything, it's that Google is a very quick study.

While Nunberg's use of "disaster" may overstate the case (and while anybody's use of "The Last Library" was always unfortunate), some of the comments also seemed a bit odd, discounting the *de facto* monopoly as nonexistent and, for several of them, basically saying "Don't like it? Don't use it."

Finding and Fixing Errors in Google's Book Catalog

Ed Felten comes at this from a slightly different angle in this [September 2, 2009 post](#) at *Freedom to Tinker* which links to Nunberg's post. He says of the post and stream of comments:

We rarely see such an open and constructive discussion of errors in large data sets, so this is an unusual opportunity to learn about how errors arise and what can be done about them.

Felten notes Google's Orwant and his lengthy comment and some of the effects of errors in this sort of catalog. But Felten's very much on Google's side, at least effectively, as in these concluding paragraphs:

What's most interesting to me is a seeming difference in mindset between critics like Nunberg on the one hand, and Google on the other. Nunberg thinks of Google's metadata catalog as a fixed product that has some (unfortunately large) number of errors, whereas Google sees the catalog as a work in progress, subject to continual improvement. Even calling Google's metadata a "catalog" seems to connote a level of completion and immutability that Google might not assert. An electronic "card catalog" can change every day—a good thing if the changes are strict improvements such as error fixes—in a way that a traditional card catalog wouldn't.

Over time, the errors Nunberg reported will be fixed, and as a side effect some errors with similar causes will be fixed too. Whether that is good enough remains to be seen.

I didn't have the impression that Nunberg regarded the metadata problems as unfixable. Rather, I believe he regarded them as so serious that Google needed to give them more attention.

The Last Library Is Greater than Google

Apparently Geoff Nunberg did indeed use the phrase "The Last Library" during the Berkeley session, based on reporting that I didn't pick up. I find that truly unfortunate. So, I think, does Barbara Fister, author of this [September 3, 2009](#) "Peer to Peer Review" column at *Library Journal*. But she'll use it for the purposes of discussion.

The terms of the settlement raise all kinds of issues. Will the Google orphanage unfairly require payments for books that are actually in the public domain? Will there be any privacy provisions for users, or will the Last Library conduct surveillance on your reading habits in order to match your interests to advertising? How expensive will it be to enter the library? Will academic libraries have to sacrifice their book budget so they can subscribe to the One and Only library? And will Google's special relationship with the class of authors and publishers mean we'll never have a second crack at building a digital library that functions differently?

She links to the *Chronicle* article (noted above) as an expansion of what Nunberg meant by "the Last Library": "He argues that Google's project to digitize books is the largest ever and, because nobody else will be in a position to do what they've done, it will essentially be the final word." Unfortunately, that assertion (quite possibly true) doesn't have quite the

same meaning; to many of us, “the Last Library” could also be taken to be “the only library left standing.” For true digital triumphalists, that might be true: After all, books on “dusty bookshelves” in libraries are anachronistic. (Note: Barbara Fister is *not* saying this; I’m riffing.)

Fister is not buying Google as the last library:

The Google library is handy when you have a precise need. It’s not much good when you’re doing what students typically do in libraries: explore an unfamiliar idea, get a sense of the landscape, and once oriented, home in on a promising area. As a friend of mine once said, they’re not seeking answers in a library; they’re learning how to ask good questions...

[I]t won’t be the last library. We will still need libraries that are more than digitized caches of information. We’ll still need places that serve a local community, that curate a collection, that organize it by both subjects and classes, making it approachable from multiple directions. We’ll still have to help students learn how to formulate questions, examine the possibilities, and gain a sense of the infinite possibilities encompassed in a library that is not infinite.

What Ms. Fister says. If only everybody believed that and acted accordingly!

Library Groups Step Up Criticism of Google Settlement; Some Academic Institutions Support It

That’s *one* title for this [September 3, 2009](#) Norman Oder report at *Library Journal*. The other one—what appeared as a title when I printed out the first page—is “GBS replace ILL?” Talk about different slants to a story!

Here’s the first paragraph—and based on that, either headline is plausible:

In a flurry of comments filed with the federal court New York overseeing the proposed Google Book Search settlement, library groups have stepped up their criticism, joined by several industry heavyweights. On the other side, a variety of supporters have emerged, notably smaller academic institutions that believe that the institutional subscription database (ISD) would be a far better deal than having to try to match a major research library. Also, one library supporter suggested that GBS could essentially replace inter-library loan.

The “library groups” (ALA, ACRL, ARL) issues are largely ones already covered. The supporting libraries are interesting. For example:

The Association of Independent California Colleges and Universities (AICCU)—also representing in this case independent colleges in Arkansas, Florida, Iowa, and South Carolina—wrote that they favored the settlement because “the cost pressures facing educational institutions have limited the ability of these traditional solutions to continue to increase content available for students and scholars.”...

Abilene Christian University endorsed the settlement, as well: “While we do not currently know the cost of this service, our expectation is that it will be significantly less than the alternatives available to us today,” the school wrote. “Without unlimited funding to purchase resources, there is truly no other way we can currently provide access to the breadth and depth of the collections in Google’s partner libraries.”

As for ILL, Oder sees a contrast between a University of Wisconsin position and an Abilene Christian position—but I’m not sure I see the contrast. Wisconsin said:

This aspect of the settlement [print-on-demand, one of the optional New Revenue Models] could also alter or eliminate the traditional interlibrary loan process. In the end, it may be more effective, in respect to both cost and time, to buy a single print copy on demand than to borrow and ship a copy from another library, resulting in additional fair compensation for the authors and publishers.

Here’s Oder’s paraphrase of Abilene Christian’s suggestion: “that the ISD would replace ILL, saving time and money as well as more clearly indicating to searchers that the material is worthwhile.”

Sounds pretty similar to me (and the suggestion that simply being a book Google scans makes it likely to be worthwhile is, I hope, not a fair representation of what Abilene Christian intended!). There’s more in an interesting news story.

A tale of 10,000,000 books

Here’s one from the Googleplex: Sergey Brin’s [October 9, 2009 post](#) on the *Google Official Blog*, which previously appeared in the *New York Times*. It’s a well-written apologia for GBS that seems to view AG, AAP and Google as The Three Amigos who had a temporary argument: “While we have had disagreements, we have a common goal—to unlock the wisdom held in the enormous number of out-of-print books, while fairly compensating the rights holders.” Brin’s out to “dispel some myths about the agreement and to share why I am proud of this undertaking.”

What are the myths?

Some have claimed that this agreement is a form of compulsory license because, as in most class action settlements, it applies to all members of the class who do not opt out by a certain date. The reality is that rights holders can at any time set pricing and access rights for their works or withdraw them from Google Books altogether. For those books whose rights holders have not yet come forward, reasonable default pricing and access policies are assumed. This allows access to the many orphan works whose owners have not yet been found and accumulates revenue for the rights holders, giving them an incentive to step forward.

It's a compulsory license with an opt-out provision. Not exactly a myth.

Others have questioned the impact of the agreement on competition, or asserted that it would limit consumer choice with respect to out-of-print books. In reality, nothing in this agreement precludes any other company or organization from pursuing their own similar effort. The agreement limits consumer choice in out-of-print books about as much as it limits consumer choice in unicorns. Today, if you want to access a typical out-of-print book, you have only one choice—fly to one of a handful of leading libraries in the country and hope to find it in the stacks.

I wish there were a hundred services with which I could easily look at such a book; it would have saved me a lot of time, and it would have spared Google a tremendous amount of effort. But despite a number of important digitization efforts to date (Google has even helped fund others, including some by the Library of Congress), none have been at a comparable scale, simply because no one else has chosen to invest the requisite resources. At least one such service will have to exist if there are ever to be one hundred.

If Google Books is successful, others will follow. And they will have an easier path: this agreement creates a books rights registry that will encourage rights holders to come forward and will provide a convenient way for other projects to obtain permissions. **While new projects will not immediately have the same rights to orphan works, the agreement will be a beacon of compromise in case of a similar lawsuit, and it will serve as a precedent for orphan works legislation, which Google has always supported and will continue to support.** [Emphasis added.]

Ahem. Brin's looking forward to lots of competitors as long as they do their own scanning...oh, and they'll have access to a registry that gives special treatment to Google and only to Google.

I'm picking nits, but it's a good statement, worth reading—and typical of the reasons why I had (and have) trouble deciding whether GBS would have been a good or a bad thing, even though I have no qualms whatsoever in agreeing that Judge Chin reached the only plausible decision.

GBS 2.0: The New Google Books (Proposed) Settlement

That's Kenneth Crews' [November 17, 2009 writeup](#) at the Columbia University Library's Copyright Advisory Office site. Crews, always articulate and worth reading, leads off with a striking paragraph:

One of the basic indicators of successful negotiations is that each party leaves equally satisfied and dissatisfied. No one gets everything. Trouble brews, however, when the deal leaves so much dissatisfaction that the good news is overwhelmed. Such may be the case with the revised Google Books settlement, offered for our consideration at midnight on November 13 ("GBS 2.0"). It is a neat deal, but the negatives are inescapable. It is hard to build an exciting new future on such [ambivalence](#).

That link is to a striking story from Frankfurt: That German publishers, upset about being included in the original GBS without being consulted, were now upset because they were excluded from GBS2.

Crews notes that GBS2 did "nothing meaningful about privacy rights" of readers—and that it still gave Google an effective monopoly on scanning and marketing of orphan works. Then there are libraries:

GBS 2.0 is a double whammy for libraries. First, the ISD's scope is slashed. No longer "worldwide," the settlement is now only about books registered with the U.S. Copyright Office (which will be dominantly U.S. books), and books originating from the United Kingdom, Canada, and Australia. Gone are all other books from Europe, Asia, Africa, South America, and other regions. Because the settlement is now tightly limited, so will be the ISD. The big and (probably) expensive database is no longer so exciting. Many of the books under GSB 2.0 are likely already available to many libraries.

The second whammy is legal. Because the settlement does not cover all books, liabilities surrounding some large portion of the books already shipped by libraries and scanned by Google are not released. Copyright owners from France, Argentina, New Zealand,

and China retain the right to commence yet another lawsuit against Google, conceivably drawing libraries into the melee. Why the libraries? Rightsholders could claim that libraries are “contributory infringers” by making the books available. Moreover, many libraries and Hathi Trust, continue to hold book scans received from Google that are now outside the settlement.

Revised Google Book Search Settlement from a Library Perspective

Peter Murray takes a look at GBS2 in [this November 18, 2009 post](#) at *Disruptive Library Technology Jester*. He doesn't include library issues that haven't changed and issues such as what really did turn out to be the killer: “the appropriateness of setting policy via class action.”

He's mostly itemizing library-related changes such as the definition of a book (for purposes of the settlement) to exclude most publications outside of four English-speaking nations; the explicit exclusion of microforms; inclusion of OCLC in institutional consortia; possible expansion of free public access terminals in public libraries. There's a slightly odd definitional issue: “book” excludes periodicals—but includes *book compilation* of periodicals, which could mean bound volumes.

Murray sees a little more improvement in privacy than Crews does:

A big part of objections from libraries is the disparity of privacy expectations between how libraries handle patron records and the more permissive way that Google logs and tracks users' activities. The amended agreement does include a new section (§6.6.f) on privacy: “in no event will Google provide personally identifiable information about end users to the Registry other than as required by law or valid legal process.” The settlement is silent on the disposition of usage records within Google. This does not satisfy the concerns of the Electronic Frontier Foundation, among others.

Murray notes that custom publishing (e.g., coursepacks and custom anthologies), which had been part of a possible new service in GBS, is gone in GBS2. He notes other changes already noted elsewhere as well. And I love this final paragraph:

Sometimes I wonder what actually goes on in some of the back-room negotiations for these agreements. For instance, according to §1.19, the definition of “Book” no longer includes calendars. Someone thought it might? Also, in the definition of “Principal Work” the example was changed from “The Old Man and the Sea” to “To Kill a Mockingbird”. A lawyer wasn't a fan of Verlag's work?

OK, GBS2 actually says “Principal Work,” not “Principle Work,” and I think the reason for the change is fairly clear: The Verlag edition of *The*

Old Man and the Sea would be excluded from the settlement, so it's not a good example. (The definition is saying that two editions of the same title that have different forewords or annotations—or even different ISBNs, as in hardcover and paperback copies—are different Books for GBS2.) But, unlike Murray, I didn't try to plow through a *long* agreement ([the redlined version](#), showing changes, is 377 PDF pages!)

Is Google Good for History?

That's Dan Cohen's question in [this January 7, 2010 post](#) at his eponymous blog; Cohen is a history professor and director of the Roy Rosenzweig Center for History and New Media at George Mason University. The post is Cohen's prepared remarks for an American Historical Association panel with the same title.

The post is just under 2,700 words. It's carefully written. I could quote the entire post here with impunity (Cohen uses a CC BY license), but it's equally easy for you to read it yourself—and the comments (there aren't really all that many: The count of 49 is mostly backlinks). It's a fairly even-handed consideration. A few excerpts, starting at the beginning:

Is Google good for history? Of course it is. We historians are searchers and sifters of evidence. Google is probably the most powerful tool in human history for doing just that. It has constructed a deceptively simple way to scan billions of documents instantaneously, and it has spent hundreds of millions of dollars of its own money to allow us to read millions of books in our pajamas. Good? How about Great?

But then we historians, like other humanities scholars, are natural-born critics. We can find fault with virtually anything. And this disposition is unsurprisingly exacerbated when a large company, consisting mostly of better-paid graduates from the other side of campus, muscles into our turf. Had Google spent hundreds of millions of dollars to build the Widener Library at Harvard, surely we would have complained about all those steps up to the front entrance...

Of course, like many others who feel a special bond with books and our cultural heritage, I wish that the Google Books project was not under the control of a private entity. For years I have called for a public project, or at least a university consortium, to scan books on the scale Google is attempting... The likelihood of a publicly funded scanning project in the age of Tea Party reactionaries is slim...

Google Books is incredibly useful, even with the flaws. Although I was trained at places with large research libraries of Google Books scale, I'm now at an institution that is far more typical of higher ed, with a mere million volumes and few rare works. At places like Mason, Google Books is a savior, enabling research that could once only be done if you got into the right places. I regularly have students discover new topics to study and write about through searches on Google Books. You can only imagine how historical researchers and all students and scholars feel in even less privileged places. Despite its flaws, it will be the the source of much historical scholarship, from around the globe, over the coming decades. It is a tremendous leveler of access to historical resources.

Google is also good for history in that it challenges age-old assumptions about the way we have done history. Before the dawn of massive digitization projects and their equally important indices, we necessarily had to pick and choose from a sea of analog documents. All of that searching and sifting we did, and the particular documents and evidence we chose to write on, were—let's admit it—prone to many errors. Read it all, we were told in graduate school. But who ever does? We sift through large archives based on intuition; occasionally we even find important evidence by sheer luck. We have sometimes made mountains out of molehills because, well, we only have time to sift through molehills, not mountains. Regardless of our technique, we always leave something out; in an analog world we have rarely been comprehensive.

In addition, listening to Google may open up new avenues of exploring the past. In my book *Equations from God* I argued that mathematics was generally considered a divine language in 1800 but was “secularized” in the nineteenth century. Part of my evidence was that mathematical treatises, which often contained religious language in the early nineteenth century, lost such language by the end of the century. By necessity, researching in the pre-Google Books era, my textual evidence was limited—I could only read a certain number of treatises and chose to focus (I'm sure this will sound familiar) on the writings of high-profile mathematicians. The vastness of Google Books for the first time presents the opportunity to do a more comprehensive scan of Victorian mathematical writing for evidence of religious language. This holds true for many historical research projects...

[C]omplaining about the quality of Google's scans distracts us from a much larger problem with Google Books. The real problem—especially for those in the digital humanities but increasingly for many others—is that Google Books is only open in the read-a-book-in-my-pajamas way. To be sure, you can download PDFs of many public domain books. But they make it difficult to download the OCR'd text from multiple public domain books—what you would need for more sophisticated historical research. And when we move beyond the public domain, Google has pushed for a troubling, restrictive regime for millions of so-called “orphan” books...

We should remember that the reason we are in a settlement now is that Google didn't have enough chutzpah to take the higher, tougher road—a direct challenge in the courts, the court of public opinion, or the Congress to the intellectual property regime that governs many books and makes them difficult to bring online, even though their authors and publishers are long gone...

That's a bit more than one-quarter of the discussion, possibly badly selected. [Go read the whole thing.](#)

Hurtling Toward the Finish Line: Should the Google Books Settlement Be Approved?

Ivy Anderson (Director of Collections, California Digital Library) asks that question in this [February 16, 2010](#) piece at CDL. She says these are “personal thoughts from my vantage point at the California Digital Library.”

CDL and indeed the UC Libraries as a whole bring what is perhaps a unique perspective to this dispute. The University of California Libraries are Google's second-largest library digitization partner; we are also the second-largest book digitization partner of the Internet Archive, thanks to generous funding in the past from Microsoft, Yahoo, the Alfred P. Sloan and Kahle/Austin foundations, and other sponsors. In all, UC Libraries have now digitized 2.5M books from their collections through these projects, both in- and out of copyright.

She notes that UC faculty (e.g., Pamela Samuelson) are “among the Settlement's most prominent critics.”

While many assume this to be an uncomfortable position, I don't find it so. Like any complex enterprise, the Google Books project is appropriately viewed from many perspectives. The proposed settlement is hardly perfect; as Google acknowledges in its brief, it's

a compromise among parties with differing agendas and motivations. CDL is a staunch supporter of the underlying aims of the Google Books project to make the knowledge enshrined in the world's great libraries discoverable and accessible across the globe, and we support the public benefits that will ensue, including the benefits to libraries, if the Settlement is approved. At the same time, public criticism has been good for the Settlement, producing very real improvements in the amended version that is now before the court; improvements that would not have been made without that criticism. Long live democracy!

She lists a few of the objections that participating libraries reasonably have over details of GBS, then concludes:

The problem with this view, of course, is that libraries did not initiate this enterprise, and we are not its only beneficiaries. The Google project placed two sets of commercial interests at loggerheads, with copyright law in the middle. Admittedly, libraries took a risk in engaging in a partnership so legally entangled.

But let's be honest: though few seem willing to admit it, revitalizing the world's heritage of books for a digital age – a task that many considered impossible only a few short years ago – appears within reach today almost entirely due to Google's enterprising vision.

She notes that CDL is a member of the Open Content Alliance (which it joined before working with GBS)—and that “when Google's competitors withdrew their support for that project, no other funders stepped in to fill the breach.” She also notes CDL's own estimate of what it would take UC to convert UC's 15 million unique books to digital form: Half a billion dollars and one and a half centuries. “And that is just the University of California's books.”

There's a lot more to this informed pro-GBS discussion, including some interesting (if not always entirely convincing) responses to some criticisms of GBS2. In the process, she includes a paragraph that hints at the underlying problem with GBS, although that's clearly not Anderson's intent:

When the [purposes that we first envisioned](#) when embarking on these projects—all arguably fair uses of this content—are reviewed against the Settlement impacts, it's hard to view the Settlement as anything but a positive development. More books will be available in full view, both to libraries and to consumers. New services will be developed for print-disabled users and for largescale

computational analysis, further unlocking digitization's transformative potential. Disclosure of rights information through a central registry (at least for U.S. books) is likely to have far-reaching impacts, facilitating the eventual orderly release of books into the public domain. Google's competitors are likely to join the push for orphan works legislation, increasing its chances of success. And with the Settlement behind us, we can all proceed in an environment of greater certainty.

And that vast expansion of scope was, plausibly, part of the problem.

What if GBS was to be rejected? She says it would "hardly be a crisis" for libraries, and that the original benefits would still be realized while fears of some objectors "will melt away like the elusive Vancouver slow." Indeed, while she argues for approval of the settlement, she appears mostly worried about a combination of rejection and "further legal setbacks" that could cause Google to abandon library digitization. Not that Google's ever been known to shut something down because it isn't going as well as they liked...

Then comes the best part, "Life Beyond Google Search," addressing concerns she's heard from UC faculty that have nothing to do with GBS as such:

To our scholars who worry that we are about to throw our physical collections overboard in favor of digital surrogates of sometimes uneven quality, I want to say: not to worry. True, libraries everywhere find themselves having to consign more and more of their physical collections to remote storage as campus space grows increasingly scarce and user preferences migrate online. And some libraries—the UCs far less than others—are addressing the space crunch by de-accessioning low-use materials that are widely held with the knowledge that they can borrow these items from another library if need be. (Many [cooperative initiatives] are now underway to share such information and ensure that enough copies are retained throughout the nation's system of libraries to protect the integrity of the scholarly record.) That train has already left the station, and it's happening independently of largescale digitization. What digitization offers is a valuable complementary mitigation strategy: we can now make those remote collections eminently browsable, saving time and expense both for users and for libraries. As a library user, you can now determine whether that book is really what you're looking for before you request it, not afterward – and in some cases, the digital surrogate may indeed be all that you need. Libraries can promote these 'hidden' volumes more effectively to their users, while

limiting delivery costs to just those items that are truly wanted. This browsable and/or searchable digital surrogate—which is the quality level that most of the Google mass digitized scans are aimed at—is not a replacement for the original print book, and was never intended to be.

To our scholars who worry that we are outsourcing our library collections and services to Google, again I want to say: please don't worry on this score either. Far from abrogating our mission as stewards of the cultural record, we who have opened up our collections to digitization are shouldering this role with vigor...[Some discussion of HathiTrust]... The digital library of the future resides not with Google, but with us. And we are building it today.

Berkeley and the rest of UC didn't plan to Throw Out All the Books. I never believed that they did. I wonder at some smaller institutions that seem to have felt otherwise.

The handful of comments are nearly all congratulatory—and this is indeed an excellent essay, one that probably deserves reading [in the original](#).

The Fight over the Google of All Libraries: An (Updated) Wired.com FAQ

With “Google of All Libraries” in the headline and “words printed on dead trees” not too far below, it's fair to assume I'm biased against Ryan Singel's [February 18, 2010 piece](#) at *Wired* as being, well, typical *Wired* coverage. But let's see what's here—starting with the lead paragraph, where Singel seems to imply that Google *always* planned the vast set of enterprises in GBS2:

Google's plan to digitize the world's books into a combination research library and bookstore started in 2002 when it first began scanning books without permission from authors. The Google Books project has since grown into an epic legal battle pitting Google and a coalition of authors and publishers who originally sued the search engine against a small army of academics, open-source advocates, Google competitors and a medley of authors.

Had you asked Google in 2002, I suspect it might have disclaimed “research library” and would almost certainly have disclaimed “bookstore,” but I could be wrong. Meanwhile, the next sentence could be a good reminder of why Judge Chin was, looking back on it, unlikely to approve GBS:

The Justice Department's antitrust division has twice weighed in against the settlement, dimming Google's chances of convincing a federal judge to **let it slice through stifling copyright law to create a vibrant online library**. [Emphasis added.]

Just at a guess, no Federal judge is likely to feel it's the judge's job to overturn decades and centuries of legislation, to "slice through stifling copyright law."

I could also question the flat statement that "Google was prepared to defend itself on [fair use] grounds before getting a better deal"; I guess we'll eventually find out. Singel is absolutely clear as to the motives of AAP and AG, and he apparently thinks that the rights of rightsholders have nothing to do with it. He says it's this simple: "Once they saw Google using snippets of the books in search results and making money off it, they decided they deserved some of it. After all, they wrote the books. At least some of them, anyway."

We're into classic *Wired* territory. Those blessed as Good Guys always have pure motives; those not so blessed are always impure. The reporting is sloppy (he says the one-free-computer-per-library is for academic as well as public libraries). He's certain that libraries will be forced to buy lots and lots of subscriptions (no matter what the price is, apparently) because demand will be so overwhelming.

It could be worse. If you're one of those who cheers "the library of the future" you might find this FAQ interesting.

Virginia Makes the Google Settlement Better for Libraries

This Peter Hirtle post [on March 21, 2010](#) at *LibraryLaw Blog* focuses on the revised agreement between the University of Virginia and Google, which made the previous agreement conform to GBS2—but also contained "two improvements over other amended agreements (such as Michigan's) that have important implications for everyone interested in the settlement."

The first important change concerns ownership of the scans of the public domain works. In the initial contracts (such as those with Michigan and California), Google retained ownership of the scans of public domain books. While Google was required to make the scans publicly available for free, libraries were somewhat limited in what they could do with those scans. For example, they could not use the scans in a print-on-demand (POD) operation, nor could they offer the scans to Google's competitors. The restrictions lasted forever... [Michigan improved this.]

Virginia's revised agreement improves on Michigan's changes. Section 4.10(d) of UVA's revised agreement stipulates that all

restrictions on the use of the public domain scans terminate after 15 years...

The second major improvement in Virginia's contract concerns the pricing of the institutional subscription. One of the great unknowns in the settlement is how much the institutional subscription will cost. Since the only comparable database—Google's own massive compilation of public domain books— is offered to the public for free, one could argue that the subscription cost of a database of copyrighted but out-of-print books should be very, very low. On the other hand, if the Books Rights Registry required Google to follow the model of some of the commercial databases of public domain and copyrighted books and newspapers—a model that tries to sell a small number of subscriptions at a very high cost—the price could be quite high. ..

The second has to do with pricing for the institutional subscriptions. You can read it in the original. I'm struck by this paragraph, however:

There is no discussion of the privacy of users of the institutional subscription, but this doesn't concern me. This is a marketing issue. If Google does not offer in the institutional subscription the protections for reader privacy that libraries demand from all of their third-party vendors, then the libraries should refuse to subscribe to the database. Since the settlement demands widespread subscriptions, Google would be forced to amend its policies. If libraries do not want to exert their market influence, they have only themselves to blame.

The word "cavalier" springs to mind and it's hard to dislodge it. The participating libraries have the most clout with Google; surely they could be expected to fight for privacy rather than saying "don't like it, don't subscribe"?

Google Book Settlement Market Analysis Q&A

Norman Oder wrote this for *Library Journal* [on April 22, 2010](#), considering these questions:

How many libraries would buy access to the millions of titles in the Google Books database, assuming the pending settlement is approved? How much might it cost? Who would market it?

The piece is mostly an interview with Michael Cairns, who prepared [a paper](#) offering his claims of the likely pricing and market penetration for the database. I won't discuss the paper directly; I'm a little bemused at the notion that 47% of public libraries (75% of which serve 25,000

people or fewer) would pay an average of \$21,000 a year for this database, and what that would do to the acquisitions budgets of any but the few hundred largest libraries.

Just for fun, I looked at 2009 data. That year, a total of 1,771 public libraries spent at least four times as much on materials as the suggested \$21,000. 2,232 (including all of those) spent at least three times as much and 2,932 at least twice as much. When you get to the 47% level, you're including every library with at least \$21,000 for *all* material acquisitions. So Cairns seems to believe that demand would be so high that hundreds of libraries would cease *all other acquisitions* in order to pay for it. What frightens me is that I'm not sure he was wrong.

In the interview or article, Cairns says that the subscription would offer "potentially great savings over interlibrary loan" and repeats his analysis claiming that there are relatively few orphan works (discussed earlier). He doesn't worry about monopolistic pricing:

I do think that Google seeks maximum exposure for the content—not only to support its stated mission of providing wide and broad access to this 'hidden' content, but also to support other business opportunities they may implement (such as advertising programs). And while I don't cover potential uses of the scanned book content to support advertising programs (or business models) these may be launched as Google rolls out the offering.

As for the \$21,000 price and assumed 47% market penetration, here's a nice way of saying "I made it up":

The 47% is a 'blended' rate in that I assumed higher or lower levels of penetration based on the size of the library. With respect to price, my price quotes are estimates of what I believe is reasonable. I've had several people who reviewed this document suggest to me that based on the expected broad and deep depth of this database, my pricing is low versus some other aggregated databases with substantially less content. On the other hand, I am sure there are some who think \$21,000 is pretty steep.

Cairns didn't think Google would handle sales directly; he thought they'd work through somebody like Gale, OCLC or EBSCO. He arrives at an estimate of \$22 per title per year in value to Google—and it's interesting to note that even with Cairns' fairly optimistic pricing and penetration, the resulting database is relatively small peanuts for Google, yielding considerably less than \$300 million per year.

129,864,880 books...or maybe not

Here's a fun pair of pieces: The first by Leonid Taycher, a Google software engineer, posted [August 5, 2010](#) on *Google Books Search*—and the second by Joe Stoker published [August 9, 2010](#) at *ars technica*.

The title of the first piece: “Books of the world, stand up and be counted! All 129,864,880 of you.”

When you are part of a company that is trying to digitize all the books in the world, the first question you often get is: “Just how many books are out there?”

Well, it all depends on what exactly you mean by a “book.” We’re not going to count what library scientists call “works,” those elusive [“distinct intellectual or artistic creations.”](#) It makes sense to consider all editions of “Hamlet” separately, as we would like to distinguish between—and scan—books containing, for example, different forewords and commentaries.

He says they like the definition of a “tome,” an “idealized bound volume.” I’d be inclined to think of this as an edition or a manifestation. (As Taycher notes, even that has problems—several pamphlets bound together by a library count as one book, while paperback and hardback versions of the same text—even typeset identically, as in a trade paperback—count as two.) He says Google’s definition is close to what ISBNs *should* represent, but there are problems with ISBNs as well. And LCCNs and Worldcat accession numbers identify “bibliographic entities” rather than books.

Then Taycher goes through the reasoning process that leads them from masses of metadata (leading to about 600 million records, or a billion with clear duplicates) down to a more likely number.

When evaluating record similarity, not all attributes are created equal. For example, when two records contain the same ISBN this is a very strong (but not absolute) signal that they describe the same book, but if they contain different ISBNs, then they definitely describe different books. We trust OCLC and LCCN number similarity slightly less, both because of the inconsistencies noted above and because these numbers do not have checksums, so catalogers have a tendency to mistype them.

We put even less trust in the “free-form” attributes such as titles, author names and publisher names. For example, are “Lecture Notes in Computer Science, Volume 1234” and “Proceedings of the 4th international symposium on Logical Foundations of Computer Science” the same book? They are indeed, but there’s no way for a

computer to know that from titles alone. We have to deal with these differences between cataloging practices all the time.

We tend to rely on publisher names, as they are cataloged, even less. While publishers are very protective of their names, catalogers are much less so. Consider two records for “At the Mountains of Madness and Other Tales of Terror” by H.P. Lovecraft, published in 1971. One claims that the book it describes has been published by Ballantine Books, another that the publisher is Beagle Books. Is this one book or two? This is a mystery, since Beagle Books is not a known publisher. Only looking at the actual cover of the book will clear this up. The book is published by Ballantine as part of “A Beagle Horror Collection”, which appears to have been mistakenly cataloged as a publisher name by a harried librarian. We also use publication years, volume numbers, and other information.

All that yields around 210 million—but that includes microforms, audio recordings, videos, maps and other stuff. Google arrives at “about 146 million” after excluding all those, and estimate that serials account for about 16 million “bound serial and government document volumes.” Leaving the final count:

After we exclude serials, we can finally count all the books in the world. There are 129,864,880 of them. At least until Sunday.

It’s fair to say *ars technica* isn’t entirely convinced, given the title of the second piece: “Google’s count of 130 million books is probably bunk.” Stoker cites the Google post and says:

It’s a large, official-sounding number, and the explanation for how Google arrived at it involves a number of acronyms and terms that will be unfamiliar to most of those who read the post. It’s also quite likely to be complete bunk.

Why? Because “GBS’s metadata collection is riddled with errors of every sort.” This leads into a discussion of Geoffrey Nunberg and the state of GBS metadata, and of course Google’s tendency to blame libraries for the errors.

It’s also the case that, aside from any library- or Google-induced metadata errors, publishers themselves can be remarkably careless about how they mark different editions of the same work. Editions of important works that can only be told apart by an examination of signature changes in their text are the stuff of bibliophile lore. And how many errors must be corrected and subtle fixes made in between printings before a “new printing” gets promoted to a “new

edition”—the answer can vary from publisher to publisher and from work to work.

In the end, this somewhat lighthearted commentary on what I believe to have been a deliberately lighthearted Google post (I’m surprised the count wasn’t “129,864,883,” as that final zero seems terribly imprecise...) becomes another valentine to Google, as in the final paragraph:

Google may not (or, rather, certainly will not) be able to solve [the metadata] problem to the satisfaction of scholars who have spent their lives wrestling with these very issues in one corner or another of the humanities. But that’s fine, because no one outside of Google really expects them to. The best the search giant can do is acknowledge and embrace the fact that it’s now the newest, most junior member of an ancient and august guild of humanists, and let its new colleagues participate in the process of fixing and maintaining its metadata archive. After all, why should Google’s engineers be attempting to do art history? Why not just focus on giving new tools to actual historians, and let them do their thing? The results of a more open, inclusive metadata curation process might never reveal how many books there really are in the world, but they would do a vastly better job of enabling scholars to work with the library that Google is building.

Always good to see the true digital triumphalists, as in Maury Markowitz’ comment responding to Geoff Nunberg’s complaints:

Bah, his complaint was precisely the out-of-date elitist arguments people make for keeping print newspapers. His primary complaint is that the “info” section was inaccurate, which it is, but that’s only an argument if anyone actually uses it. I don’t, and GBS is my primary source of information of all sorts. Fixing this is polishing the brass on the Titanic.

I trust Google’s count many, many times more than I trust old-tech systems. At least Google can directly compare text, for instance. THAT will let you know if two books are the same, and the only reason we didn’t do this before is because we couldn’t.

To which another commenter responds (in part):

It’s hardly elitist to request accurate metadata, such as publishing date and author, and genre categorization which is not completely misleading and/or geared toward B&N stores. While massively erroneous data may not be problematic for identifying and working

with a single text, it becomes extremely problematic if one wants to do mass computation or analyze the body of work.

Ah, but Markowitz doesn't use that data, therefore that data is useless. Get it?

Reading both stories, I'm guessing that a reasonable estimate of the number of books Google might eventually be able to lay its hands on, using Google's own definition of "book," is probably somewhere between 116 million and 143 million. Or not. And that nobody in Google really believes any number past the "1" is particularly reliable—but I could be wrong on that.

The trouble with Google Books

This piece, [published September 9, 2010](#) on *Salon* by Laura Miller, may seem a little belated, as its subtitle—"How rampant errors threaten the scholarly mission of the vast digital library"—harks back to Nunberg and the metadata flap. The second paragraph recalls Nunberg's *CHE* article from 2009. A bit later, Miller notes that "much of the incorrect information remains in place."

Turns out the article is a somewhat belated interview with Nunberg, who uses his pet phrase "last library" at least twice in the discussion. He adds some tidbits I hadn't seen before, such as this (in discussing one of the more egregious errors, tagging Henry James as the author of *Madame Bovary*):

I thought it was a machine error, too, but Google assured me that they had people doing this by hand. In some cases, they got their metadata from a provider in Armenia. They say that they want to have a diversity of sources to get a more complete classification for every book, but that's just silly. The metadata at the Harvard Library was done by hand by smart people who know how to catalog.

People at Google are also saying, "Let's crowdsource this," but that is a stupid idea. You and I are both smart, knowledgeable people, but I wouldn't trust either of us to do the skilled work of cataloging a 1890 edition of "Madame Bovary." It's very difficult. It has to be coordinated by uniform standards. An example of the kind of mess you get when you don't use uniform standards is Wiktionary (the lexical counterpart of Wikipedia). Unlike an encyclopedia, a dictionary isn't useful unless it's consistent in style. And metadata is hard to fix if you don't get it right in the first place. Someone has to spend a lot of money to properly catalog a research library, and I don't know if Google understood that going into it.

I don't remember Nunberg previously saying not only that Google Books is the "last library" but also that it's a *research library*. He also discusses some of the other problems with the scanned books.

Interesting discussion, even if it largely covers ground already covered.

The Illusion of Google's Limitless Library

This "Library Babel Fish" column by Barbara Fister [on December 4, 2010](#) at *Inside Higher Ed* is the last pre-decision item I tagged relating to libraries and metadata. The column was triggered by Google's launch of what's now called Google Play, its ebook retail and reading platform, and the [cute video](#) it did (it was called Google eBooks at the time).

It's all about choice! You can choose any book and read it anywhere on any device! Sounds pretty sweet.

It's certainly true that the video says "it's all about choice" more than once, that it says *your library* will be stored in Google's cloud, *all your books* (apparently print books became wholly irrelevant at this point), and indeed "read it anywhere on any device" (although "almost" might have crept in there).

But it's not *quite* true. Google would like us to think that they've digitized every book, that any book ever published that you may want to read can be plucked from the cloud and read anywhere.

The video doesn't explicitly make that claim. It does say "millions of books"; it doesn't say *every* book (except implicitly, when it says *it* will house *your entire library*, no matter who you are. That's a pretty broad claim. And, as Fister demonstrates in the rest of the column, it's not even close to being true.

The rest of the column deals with that and ancillary issues. It's a good one. The comments are interesting.

With Google Settlement Rejected, Library Groups Keep Eye on Access

And then it was over—at least the GBS phase. You'll see more about what happens after Judge Chin's decision in the final section of this roundup, but this piece and the next seem primarily focused on library issues and make more sense here.

Josh Hadro wrote this [on March 24, 2011](#) at *Library Journal*. He notes the background for librarians and libraries:

What the vast majority of librarians hoped to see out of this lawsuit was a precedent-setting determination on the fair-use right to index and search copyrighted materials (recalling the scope of the initial complaint against Google). Barring that, most considered an acceptable consolation prize to be easy access to a full-text union

archive of the nation's premier research collections, as the settlement would have provided.

As of Tuesday, neither of those options are in the offing. What librarians can look forward to instead: a renewed commitment from library advocates to make more content accessible to scholars and to the general public, whether via an alternative settlement agreement or legislative recourse.

Hadro discusses possible next steps and links to "[GBS March Madness](#)," a remarkable flowchart (originally from 2010, but with a new circle saying "You Are Here"). He cites three options: An appeal of Chin's decision, a restructured settlement or resumption of the long-delayed trial. He says most sources thought the likelihood was in that order, but that James Grimmelmenn believes a revised settlement was most appealing.

Hadro quotes some key reactions and you're better off reading them at *LJ*. With regard to the orphan works issue, now back in the hands of legislators, he concludes:

While librarians recognize that legislative attention to the orphan works issue has the greatest potential benefits, many are wary of the long timelines involved in such endeavors and the very real danger of opposition from any number of industry interests.

Piling on

Finally for this section, here's Kevin Smith [on March 26, 2011](#) at *Scholarly Communications @ Duke*. Either Smith follows a very different set of commentators than Hadro, or the sense of "most sources" changed rapidly:

I have been interested to see that no one else whose comments I have seen seems to think that an appeal is likely. Indeed, I draw that conclusion entirely from the absolute silence I find about that option, while there is much discussion of other possibilities.

I imagine the reason for this is the strong sense that the rejection was, as Prof. Pamela Samuelson puts it in [this interview](#), the only conceivable ruling that the judge could have made and that it is quite water-tight from a legal perspective. While it is not unheard of for parties to spend lots of money on lost causes, the majority of commentators obviously feel that Google, the Author's Guild and the Association of American Publishers will not throw good money after bad by filing an appeal.

The next sentence suggests that "most sources" might have been Kevin Smith in the first place: "I am perfectly willing to pile on to this bandwagon, abandon my speculation about an appeal, and think about

what other options the rejection might open up.” And he notes the critical need: librarians and others need to reengage the orphan works issue. But how to do that?

The Google Books Settlement gave librarians, copyright activists and even Congress a chance to sit back and assume that orphan works was being dealt with. Sure, we thought, there are millions of works that are still protected by copyright but for which no rights holder can be found; access to these works is a problem, but Google is going to solve it. Now we cannot look to Google for a solution, so it is worth revisiting what a sensible solution might look like.

I think we should consider the possibility that a legislative solution may not be either the most practical or the most desirable way to resolve the issue of access to orphan works. The orphan works bill that came closest to passing a few years ago was hardly ideal, since it would have created requirements both burdensome and vague for gaining a measure of extra protection from copyright liability. A good bill that really addresses the orphan works problem is probably both hard to conceive and unlikely to pass. So what alternatives short of a legislative solution might we consider?

The obvious answer is fair use, since most proposals for orphan works solutions would essentially codify a fair use analysis. Fair use, after all, is really an assessment of risk, since its goal is to reuse content in a way that wards off litigation. The Congressional proposals around orphan works would have simply reduced the damages available in defined situations, thus also having as a primary purpose the reduction of the risk of litigation. Careful thinking about projects like mass digitization of orphan works can accomplish the same goal by balancing analysis of the public domain, permissions where they are possible and needed, and a recognition that for truly orphan works, the fair use argument is much stronger since there is no market that can be harmed by the reuse.

The rest of the piece discusses how orphan works might be dealt with on a fair use basis. You’d need to read it in the original. Could it plausibly work in the absence of legislation? Well, Smith is a lawyer and many times as knowledgeable about copyright as I could ever be, so I’ll just [point you to his article](#).

Authors and Publishers

Just a few items here from the specific perspectives of authors (not necessarily the Authors Guild) and publishers.

Google Book Search Settlement: A Publisher's Viewpoint

This piece is an interview of Oxford University Press's president (Tim Barton) and general counsel (Barbara Cohen) by Mary Minow, appearing [in September 2009](#) at *Copyright & Fair Use*, a Stanford University Libraries site. If you're a regular *C&I* reader, you may be aware of Oxford University Press (OUP) as one of three plaintiffs in the Georgia State University case—a university agency that sues other universities on behalf of copyright maximalism. This interview, of course, was on an entirely different topic.

OUP, not an AAP member, came out in support of the settlement. Says Barton (in part):

When we did understand [the settlement], what made it in the end straightforward for us to support the settlement was the almost unimaginable access that it will enable to millions of works that were lost to readers and scholars and which, without the settlement, were likely to remain so. We had been working on a project at OUP to bring our own out-of-print books back to life, and we were aware of the very considerable difficulties and costs involved in doing so. From these efforts at digitizing our backlist, we saw that only an entity such as Google would take on the risks and make the investments needed to bring these millions of books back to life. This is because Google wants to make its search engine as useful as possible, in order to secure advertising revenues, and so it can justify the major costs: publishers cannot make anything like the same level of return on selling their out-of-print backlist as Google can in securing revenues as a result of returning the best quality searches.

After this love letter to Google, Barton comes out swinging at settlement opponents in explaining why OUP publicly supported a settlement that it wanted improved (in ways that aren't explained):

We decided that we should publicly voice our support for a number of reasons, including what I view as poor branding of this settlement as “the Google settlement.” It is not surprising that the public has been especially cautious—skeptical even—in considering something that sounds as if it is just for the benefit of a company as powerful as Google. But this isn't just Google's settlement; Google is a party to the settlement, for sure, but it is equally a settlement which is in the

interests of publishers, authors, libraries, and, I believe, the general public. We also felt that while the groups that had negotiated the settlement had done a remarkable job in negotiating it, they were falling short in explaining and promoting it. Those who had negotiated the deal didn't seem to be coming forward to correct misunderstandings and support it. I can appreciate that, after having slogged through two and a half years of negotiation, they must have relished the prospect of putting it to the side even for a short while. But the vacuum created was filled by outspoken critics, some of whom seemed to have vested interests in scuttling the settlement. Underlying a growing chorus of criticism, we heard repeated misunderstandings about the settlement, as well as a visceral fear of something that seemed to be for Google. But, as I mention above, the settlement was negotiated by authors, publishers and libraries too, and it promises tangible and significant benefits for these groups as well.

I'm not one who eagerly sticks his head above the parapet, but I was quite concerned that, if people did not step forward to voice support for the settlement, it might fail. And that would serve no one except Google's competitors.

The settlement was negotiated by "libraries too"? That's the first I've heard of libraries being involved in settlement negotiations.

The discussion turns to antitrust, where Barton disclaims special knowledge—and then Barton poo-poops the possibility that the institutional subscription price would be too high. He makes much of that Single Free Terminal in public libraries as assuring that Google (and its publisher partners) won't gouge. You can read that argument yourself. I have to balance it against publisher records in charging just a little more than the market will bear for library subscriptions...

Regarding orphan works, there's a particularly revealing statement:

But an imperfection I see relating to orphan works is that, at least immediately following the settlement, Google alone has the ability to exploit orphan works, when even the original publishers of these works will share no such right.

If a publisher *has* rights in an out-of-print book, it's not an orphan. If those rights have reverted to the author, the publisher *should not* have special rights at that point. If publishers don't keep track of which contracts have reversion clauses and which don't, that's a different issue.

There's more here, and a couple of paragraphs later Barton repeats his claim that libraries were directly involved in the GBS negotiations:

The interests of libraries, too, seem to have been well represented—no surprise, as they were involved in the Google Library Project from the start and were at the negotiating table.

Why have I never heard this from anybody else? Why do I doubt it?

Amazon Accuses Someone Else of Monopolizing Bookselling

This item is an institutional statement from the Authors Guild, [posted September 2, 2009](#) on the AG website, and says AG is “compelled to state the obvious” after Amazon filed a brief opposing GBS:

Amazon’s hypocrisy is breathtaking. It dominates online bookselling and the fledgling e-book industry. At this moment it’s trying to cement its control of the e-book industry by routinely selling e-books at a loss. It won’t do that forever, of course. Eventually, when enough readers are locked in to its Kindle, everyone in the industry expects Amazon to squeeze publishers and authors. The results could be devastating for the economics of authorship.

After explaining that GBS is “about out-of-print books” (not quite true), AG says “Google would get no exclusive rights under the agreement” (also not quite true) and concludes “The public has an overwhelming interest in having the settlement approved.”

No further comment.

It’s interesting that the AG page devoted to the settlement includes *nothing* that directly reflects the rejection of the settlement. Just not there—although under “Press Resources” there’s a link to the [Settlement Website](#) (*not* an AG site), which includes a two-sentence statement on the rejection.

Academic Author Objections to the Google Book Search Settlement

That’s the title of Pamela Samuelson’s paper, published in the *Journal of Telecommunications and High Technology Law* and deposited in SSRN [on February 16, 2010](#). It’s a 29-page PDF. Here’s the abstract:

This Article explains the genesis of the Google Book Search (GBS) project and the copyright infringement lawsuit challenging it that the litigants now wish to settle with a comprehensive restructuring of the market for digital books. At first blush, the settlement seems to be a win-win-win, as it will make millions of books more available to the public, result in new streams of revenues for authors and publishers, and give Google a chance to recoup its investment in scanning millions of books. Notwithstanding these benefits, a closer examination of the fine details of the proposed GBS settlement

should give academic authors some pause. The interests of academic authors were not adequately represented during the negotiations that yielded the proposed settlement. Especially troublesome are provisions in the proposed settlement are the lack of meaningful constraints on the pricing of institutional subscriptions and the plan for disposing of revenues derived from the commercialization of “orphan” and other unclaimed books. The Article also raises concerns about whether the parties’ professed aspirations for GBS to be a universal digital library are being undermined by their own withdrawals of books from the regime the settlement would establish. Finally, the Article suggests changes that should be made to the proposed settlement to make it fair, reasonable, and adequate to the academic authors whose works make up a substantial proportion of the GBS corpus. Even with these modifications, however, there are serious questions about whether the class defined in the PASA can be certified consistent with Rule 23, whether the settlement is otherwise compliant with Rule 23, whether the settlement is consistent with the antitrust laws, and whether approval of this settlement is an appropriate exercise of judicial power.

The article is actually 21 pages long (roughly half that space taken up with more than 120 footnotes), followed by a list of academic authors who objected to GBS. It’s an interesting read, one that explicitly says GBS would *not* have resulted in a library, but rather a commercial enterprise. Worth reading—and a good precursor to the final decision.

Thousands of authors opt out of Google book settlement

This last piece is by Alison Flood, [posted February 23, 2010](#) at *The Guardian*. The subtitle (or deck or whatever it’s called):

Some 6,500 writers, from Thomas Pynchon to Jeffrey Archer, have opted out of Google’s controversial plan to digitise millions of books

That’s the core of the story, noting some of the better-known authors who opted out (in some cases their estates opted out). It quotes a few authors, including this surprising comment from Gwyneth Jones (who opted out “on the advice of my agency”):

Then I was inspired to read the small print too, and I didn’t like what I found. Google’s preemptive action has ‘turned copyright law on its head’. It seems they plan, unilaterally, to take ownership away from the

writer, and the ownership doesn't pass to the readers (fat chance!) but to a giant profit-making corporation.

Take ownership away from the writer: That's one of the more extreme readings of GBS. It's also interesting because Jones is one who makes most of her recent novels available for free online in "portable document format" and says "they do my sales no harm at all." Oh, and the works available through GBS would cost them "effectively nothing at their point of entry," which is an interesting economic analysis of the \$125 million settlement plan.

Class and Standing

Now we turn to the situation *after* Judge Chin rejected GBS2 (although the section following this one includes one earlier item). Most items here are by James Grimmelmenn posting at *The Laboratorium*, and all items in this section appeared within the last eight months (that is, the earliest is from December 22, 2011). It's interesting to see the extent to which Google is willing to at least indirectly contradict itself, now that the settlement is off the table.

Google Moves to Dismiss

Grimmelmenn leads off with a great first paragraph in [this December 22, 2011](#) *The Laboratorium* item:

Google gave me an early Christmas present today: a motion to dismiss the Authors Guild as a plaintiff from its case against Google, with plenty of interesting legal details to unwrap. The motion also seeks to have the American Society of Media Photographers and other visual artists' groups dropped from their own suit against Google. If granted, the motion would leave behind only individual artists and authors in the two lawsuits against Google. We have the full motion and supporting materials [at the Public Index](#); all of the legal argument is in [the supporting brief](#).

This is all about associational standing—whether AG and other groups can legitimately sue Google.

The reason why not is simple. Ordinarily, only the person who has allegedly been injured by the defendant's actions has "standing" to sue. If I'm outraged at the putrid food you were served at Burger Lord, I can't just rush off to court to sue Burger Lord. It's your call whether to sue them, not mine, and if you do, you need to direct the lawsuit yourself. My outrage doesn't give me standing; your food poisoning does.

The associational exception is how the Sierra Club can sue: Because its members have standing. Google argues *now* that associational standing is inappropriate in this case, given that (Google believes) a judgment on infringement requires deciding both whether the author actually holds rights for electronic publishing and whether Google has a fair use defense for that book. And then Grimmelmann starts having so much fun that I *must* quote directly:

Google's arguments on both of these points are interesting. When it comes to ownership, Google's brief effectively asserts that the e-rights situation for books is a tarpit in a bog under a swamp shrouded in fog. You want to sue us as an association, it asks? Fine. Just sort out who owns e-rights throughout the publishing industry first. Get back to us when you're done. (In one especially clever bit of lawyering, Google quotes guides published by the Authors Guild and the ASMP to make its point that book licensing is complicated and messy.)

As for fair use, it helps to think of this motion as a trailer for Google's opposition to class certification, coming to a courthouse near you in January. Google argues that the "individualized analysis" required by fair use will vary extensively from book to book and artwork to artwork. Some books are creative; some are more informational. Some are in print; some are out of print. Expect to see a more detailed version of this argument rolled out in January, when Google argues that the class of plaintiff authors is simply too diverse to litigate as a group.

Good stuff.

Who Speaks for Copyright Owners?

This [December 30, 2011 post](#) at *The Laboratorium* grows out of comments on the previous post (above). It's fairly long and probably worth reading in the original. I love one early statement, given the history of the case:

In part, this is due to the usual skepticism that arises whenever lawyers make apparently inconsistent arguments. "The vase was either already broken when you gave it to me, or I returned it in perfect condition" never goes over well—even though it may be the safest response a lawyer can give before the other side has informed her *which* vase her client is accused of breaking.

Setting aside Google's inconsistent arguments, he notes that the substantive issue is "what is legal to do with books?" and that the *procedural* question boils down to the post's title, here reworded as "who

is entitled to speak for copyright owners?” He finds one answer—“only individual copyright owners can speak for themselves”—unsatisfactory for at least three reasons:

First, we *don't know* what the law is on all sorts of issues. A copyright statute written without computers clearly in mind, and with digitization clearly not in mind, simply doesn't explain clearly what the rights of owners and readers are in many cases. Fair use's boundary is intentionally fuzzy; the scope of library rights has become unintentionally so...

Second, as the frequently dismissed prospect of individual author lawsuits against Google demonstrates, individual owners would be at a severe disadvantage trying to sort out their rights if they were entirely on their own. They need *something* better than the right to sue Google *seriatim*, each running up million-dollar legal bills. That something could be class actions, it could be suits by publishers who hold large portfolios of rights, it could be setting precedents that other copyright owners could use—but there has to be *something*. The possibility of mass infringement requires some possibility of mass response.

And third, in some cases, too many widely dispersed rights lead to chaos, confusion, and impoverishment. The classic parable here is that if an airplane at 35,000 feet trespassed on each house it flew over, air travel would be impossible...

Grimmelmann now says that GBS was “an extreme example” and “far too broad a delegation.”

The class action certification that the Authors Guild now seeks would let it speak on behalf of all book authors to stop Google's book scanning. And that's actually more or less how the HathiTrust suit against the library partners works, as well. Even though that's not a class action, the injunction it seeks isn't limited to the handful of authors who're suing. Instead, the Authors Guild wants the entire HathiTrust database impounded so no one can access it. In effect, it's asking, on behalf of all copyright owners, to have the HathiTrust database shut down. Some of them are presumably eager to have their works included.

There's more here; well worth reading.

Academics Object to Class Certification in Google Books Case

It's not just Google who thinks AG shouldn't be able to claim class standing, as [this February 16, 2012](#) David Rapp article at *Library Journal's*

“The Digital Shift” makes clear. Yes, we’re back to Pamela Samuelson, this time with a [letter](#) signed by more than 80 academics “asserting that academic authors should not be included as part of a class authorization.” The letter also makes an interesting claim:

We believe that our works of scholarship are more typical of the contents of research library collections than works of the three named plaintiffs in this case. Betty Miles is the author of numerous children’s books. Jim Bouton is a former baseball pitcher who has written both fiction and nonfiction books based on his experiences as a baseball player. Joseph Goulden is a professional writer who has written a number of nonfiction books on a variety of subjects, including a book about “superlawyers.” None of these three are academic authors. Their books are aimed at a popular, rather than an academic, audience. As professional writers, their motivations and interests in having their books published would understandably be different, and likely more commercial, than those of academic scholars. Hence, our concern is that these three do not share the academic interests that are typical of authors of books in research library collections. As we explain further below, the clearest indication that the named plaintiffs do not share the same priorities typical of academic authors is their insistence on pursuing this litigation.

One more quote (also quoted in full by Rapp):

It bears mentioning that despite our having raised numerous objections and concerns about the proposed settlement in a very public way by putting them in the court record, none of us has been contacted by the proposed class representatives, the Authors Guild, or the lawyers who want to be designated as class counsel to ask for our opinion about what our interests are, whether to pursue this litigation, what relief to seek, on what terms to settle it, or anything else.

“We represent you. We have no interest in what you have to say.” Sounds right to me.

GBS: A Matter of Standing and The Class Certification Fight

Back to James Grimmelmann and *The Laboratorium* for two posts, the first (“A Matter of Standing”) [on February 26, 2012](#), the second (“The Class Certification Fight”) [on March 4, 2012](#).. As with other Grimmelmann posts, I *could* legally quote the entire posts (he writes under a CC BY license), but that seems like overkill and a distraction,

since Grimmelmann's an eloquent writer and gets some interesting comments.

The first post deals with arguments raised in court hearings on standing in the Google Books and HathiTrust suits. First, Google says that the Copyright Act *specifically* prohibits associational standing. "That's a categorical argument. It would apply to all copyright cases." Google also offers more particular claims related to fair use and ownership.

Regarding fair use and Google's claim that it raises factual issues that vary from book to book, Grimmelmann says:

I'm skeptical of this objection—even if the fair-use case varies from book to book, it's quite possible that some broad lumping (e.g. books in print and books out of print) will suffice. You don't necessarily need to bring every author individually into court to decide whether, say, snippet display of fiction is or isn't fair use.

Regarding ownership, Grimmelmann notes that Google's [brief](#) does a nice job of using the associations' own copyright guides to illustrate how complex book copyright licensing is and finds this the "best-argued part of Google's briefs." I'll omit the HathiTrust discussion; that's a different (although definitely related) case.

The second post relates to certification rather than associational standing. Here's AG's group that it asserts it should be certified to represent:

All persons residing in the United States who hold a United States copyright interest in one or more Books reproduced by Google as part of its Library Project, who are either (a) natural persons who are authors of such Books or (b) natural persons, family trusts or sole proprietorships who are heirs, successors in interest or assigns of such authors. "Books" means each full-length book published in the English language and registered with the United States Copyright Office within three months after its first publication.

AG sure as hell doesn't represent *my* interests. Nor, apparently, does it represent at least 80 academic authors (see the preceding story). Grimmelmann notes why class certification is so important:

While the lawsuit could in theory go forward even without the class, it would be far less viable in practice. The prospect of a huge financial recovery both gives the Authors Guild more leverage against Google and makes its lawyers more willing to work on a contingency basis. So fighting class certification is a no-lose proposition for Google: in the best case, the case goes away, and in the worst, it would still have to litigate the fair use issue anyway.

Google did something “supremely clever”: it paid an expert \$100,000 to survey authors:

The survey shows that fifty-eight percent of authors affirmatively approve of the inclusion of their books in snippet view; fourteen percent affirmatively oppose that inclusion; and twenty-eight percent neither approve nor disapprove. *Id.* at 14. Forty-five percent believe inclusion in snippet view helps sales of their books; four percent believe it harms those sales; and fifty-one percent believe it has no effect one way or the other. *Id.* Nineteen percent believe inclusion in snippet view advances their economic interests more generally; eight percent believe it harms those interests; and seventy-four percent believe it has no effect one way or the other. *Id.*

So *most authors* when asked disagree with AG. Interesting.

Court Says Authors Guild Has Standing To Sue Over Google Books, Despite It Not Representing Authors' Views

If you think the title of this Mike Masnick post [on June 1, 2012](#) at *techdirt* is a trifle snarky, there's also this just below it: “from the *unfortunate* dept.” Gee, Mike how do you *really* feel about this?

He notes Judge Chin's reasoning and then argues with it:

If the court is going to lump different groups of authors into different camps, then shouldn't each of those groups create their own class action suits, rather than putting them all under the Authors Guild's umbrella? No one is arguing that there can't be a class action lawsuit if the relevant class is assembled. There's just a big question over whether or not the Authors Guild really represents the interests of the people included in the classes. And the judge doesn't really address that question, other than to say those who don't have a problem with Google Books can more or less opt-out of the class.

And adds:

On one point, however, the judge's reasoning does make sense: why did Google wait so long to challenge the Authors Guild's standing. Elements of this case have been going on for many, many years. It does seem a little off to file this particular point so late.

Well, see, Google didn't want to challenge standing when that standing would have been to Google's advantage... Mesnick continues:

In the long run, I still think any result only ends up harming the Authors Guild. They are showing themselves to be anti-innovation

luddites who disregard the interests of the majority of their members, while grandstanding against any new technology that upends the old publisher-gatekeeper model. That may be useful for some big name authors they represent—since it's all about keeping out competition from new authors, but it's no path to the future.

Wow. Luddites, grandstanding, keeping out new authors, and disregarding the majority of *their members*. Who knew? (A comment attempts to support that “majority of their members” by citing the Google survey of authors—but that survey was of 800 authors, not 800 Authors Guild members. Still, Masnick—in the comment stream—uses this survey of 800 authors who weren't even asked about AG membership as direct support of a claim that a majority of Authors Guild members support GBS.)

Three More from Grimmelmann

That Masnick post is a little out of chronological order. Filling in the pieces, here are three James Grimmelmann posts at *The Laboratorium*: “GBS: To Certify a Class” [on April 4, 2012](#); “GBS: Oral Argument Recap” [on May 4, 2012](#); and “Google Books Class Certified” [on May 31, 2012](#). (If Grimmelmann deliberately dropped “GBS” at that point, I think he's right: It was no longer about GBS, since that was no longer on the table.)

In the first piece, Grimmelmann admits he finds the arguments for and against class certification “a little anticlimactic.” He discusses three objections to certification:

- **Unrepresentative Plaintiffs:** Here Google notes its survey (which started with 142,000 published authors, tried to reach 10,000 of them and 5,000 by email—and ultimately got 880 responses) as saying that most authors wouldn't favor the lawsuit—and the plaintiffs say the survey's flawed and irrelevant. Oh, and those academic authors? Well, one of the plaintiffs is the widow of an academic, and she knows academic authors all care about this stuff, so... (Grimmelmann says “both nonresponsive and patronizing,” which seems about right.)
- **Copyright Ownership:** Here, Grimmelmann *does* find the response convincing—that Google's claimed complexities won't undermine the lawsuit.
- **Fair Use:** Google offered a few examples of what seems likely to be its fair use defense—the distinctions among books that make an overall decision unfair—and Grimmelmann isn't impressed: “These arguments increasingly strike me as small beer.” Go back [to the article itself](#) to see the examples and more (much more) of what Grimmelmann has to say. Then again, this is a “curse on both your houses” situation, as he says this about the plaintiffs' response:

The plaintiffs' reply here is ... interesting. The five pages in which they discuss fair use and common questions are partly an argument that these supposedly fact-specific questions can indeed be resolved on a class-wide basis. But much of the discussion is taken up with arguments on the merits: that what Google is doing is categorically, across the board, unfair. "There are no true individual questions here," seems to be the message, "because the case against Google is so overwhelming in each and every individual case."

On balance, Grimmelmann found himself "more sympathetic to the class certification motion than I expected to be."

As to the second item (the recap of oral arguments), I'm mostly pointing you [to the post](#) as an interesting summary of what happened. I'll quote his general observations, deliberately separated from the rest of his comments by a horizontal rule (names are of people who argued during the hearing):

A few general observations. First, Judge Chin's questions were thoughtful. He wasn't trying to press the parties on their weak spots; his questions were clearly directed to clarifying where the key areas of dispute were. Second, at least from the perspective of someone who wasn't in the courtroom, the case was well-argued on both sides. Zack and McGuire seem to have a slightly easier case on these motions, and they extracted some concessions from Durie with Judge Chin's help. But for her own part, she made some good points: the subtle but well-argued kind that one would expect from a real pro.

Third, the parties danced a bit around one of the key questions: what, precisely, is the allegedly infringing conduct for which the Authors Guild seeks to hold Google liable. Durie suggested at one point that the "right" at issue is the right to display a small excerpt of a book. Zack didn't reply directly, but in other briefs and arguments, the Authors Guild has framed the case as being about the mass scanning, the distribution of copies to libraries, and the security risks of holding a complete corpus. This is presumably going to be sorted out sooner or later, quite possibly by Judge Chin himself.

It's hard to predict what will happen next. My uninformed read is that today was a tactical victory for the plaintiffs: Google didn't offer a compelling argument for why the case can't proceed as a collective lawsuit. But that may not be strategically significant: the case is clearly heading towards the real battle over fair use, and I didn't get the sense

that the Authors Guild significantly improved its position in terms of selling Judge Chin on its claim that scanning and indexing is unfair. That may just indicate that Judge Chin, quite properly, is focused on the procedural motions currently in front of him. Or it could be a sign that the Authors Guild doesn't have enough arrows in its quiver to hit the no-fair-use target.

Stay tuned ...

The third post recounts Judge Chin's "eminently pragmatic" decision to allow the Authors Guild to represent its members—and to certify a class consisting of all authors with books scanned by Google. (He did the same for ASMP, the American Society of Media Photographers, for a parallel lawsuit which, among other things, deals with the *covers* of many of those books.)

This story is interesting for several reasons. Grimmelmann thinks that the opinion should worry Google a little, and elaborates on the pragmatism and how it might benefit Google as well:

Yes, some authors will have assigned away their complete copyright interests, retaining no royalty rights, and therefore will not be "beneficial owners" with standing to sue. But it will be much easier to ask authors to produce their contracts to show that their books are included in the class than to force them to sue Google individually. This portion of the opinion offers Google its best news of the day, I think: the company could throw some serious sand into the class action gears by making thousands or millions of authors pull their contracts out of the closet.

As for fair use, Grimmelmann quotes directly from the opinion:

While different classes of works may require different treatment for the purposes of "fair use," the fair-use analysis does not require individual participation of association members. The differences that Google highlights may be accommodated by grouping association members and their respective works into subgroups. For example, in the Authors Guild action, the Court could create subgroups for fiction, non-fiction, poetry, and cookbooks. In the ASMP action, it could separate photographs from illustrations. The Court could effectively assess the merits of the fair-use defense with respect to each of these categories without conducting an evaluation of each individual work. In light of the commonalities among large groups of works, individualized analysis would be unnecessarily burdensome and duplicative.

He comments “makes sense to me.” Then follows with more material that he thinks Google should find worrisome. And concludes:

This is not at all a decision on the merits. But it is still a very big deal, because it means that there will be a decision on the merits. The case is now definitively headed towards the gigantic fair use showdown everyone expected when it was filed in 2005. Google remains confident of its fair use case, I am sure, as the Authors Guild remains confident of its no-fair-use case. In the next few months, we will see the details.

Point to the plaintiffs.

The Future

This last group is mostly items that look to the future of the Google (and related) lawsuits after Judge Chin’s rejection of GBS—except for the first, which assumed that GBS *would* be approved.

5 Ways The Google Book Settlement Will Change the Future of Reading

This moderately long (for online) story by Annalee Newitz appeared [April 2, 2010](#) at *io9*. The author says the story breaks down all the complexities of GBS and “the future of books,” says GBS “could easily be the twenty-first century’s most important shift in how we deal with copyright in the world of publishing” and provides a little backstory about the Copyright Term Extension Act (the Sonny Bono Act), claiming that that act “gave birth to a loosely-organized but powerful movement of copyright reformists.” I’d suggest that’s wrong on two counts: Copyright reformers were around *long* before 1998—and for a “powerful movement” it’s been astonishingly lacking in accomplishments. Somehow, though, this copyright reform leads to GBS:

One of the basic injunctions of copyright reform is “share your culture,” and the seeds of the GBS come from an admirable Google project aimed at sharing the knowledge from research libraries with the world.

The next paragraph, in noting what Google Book Search would originally have done in terms of full-text searching and snippet views, makes this statement that *simply does not follow*: “The Mickey Mouse Protection Act may have stalled the growth of the public domain, but the company’s Google Book Search project would broaden it.” No, sorry, but making books searchable does not place them in the public domain.

Interestingly, although Newitz says that GBS had not yet been approved and might be revised, she also says flatly:

That said, the GBS *will* ultimately “turn copyright on its head,” as critics like Ursula Le Guin have said. And that will change the way you find and read books. Here’s how.

Um, no. Newitz at that point is providing 100% assurance that GBS would be approved; otherwise, the emphasized “will” is nonsense.

Then come the five ways GBS would have “changed the future of reading,” and what an odd lot they are! Boldface sentences from the headings; my notes (such as they are) in normal type.

➤ **It may become harder to get information online about books from writers you love.** Huh? Well, see, thousands of authors opted out. (Mandatory Le Guin quote follows.) The argument here is that books by those authors will be “increasingly hard for people to learn about” because, I guess, people will ignore any books that aren’t wholly readable in Google Books. There’s more here and the discussion’s complicated or confusing enough that I’ll need to refer you back to the original. It’s a little too bizarre for me.

➤ **You will find yourself reading free books online, by authors who have disappeared. And Google will make money when you do.** This part’s a little clearer (and Newitz says there are “at least a million” orphan works, without supporting evidence) but I wonder about some of the details offered.

➤ **Google will be competing with Apple and Amazon and everybody else to be your favorite online bookseller.** A long discussion that is generally reasonably sound. But then:

➤ **Libraries and bookstores will be the same thing.** Followed by “Ultimately what Google has done is transform libraries into bookstores.” **Bullshit.** I’m sorry, but that’s just plain bullshit, and the explanation doesn’t help—especially because she doesn’t even attempt to justify that absurd overstatement, but goes on to claim that GBS “regulates libraries” and to discuss privacy issues. Those issues are relevant, but tainted by the nonsense introduction.

➤ **Pulp science fiction will make a comeback in ways you might not expect.** We get more of the author’s idea of hybrid library/storefronts “whose job it is to preserve and monetize books”; there is *no* apparent possibility that public libraries might continue to own and circulate *actual physical books*; nope, libraries are all now just library/bookstores peddling Google’s goods. And this is *great* because it means “more pulp fiction, or cheaply-produced and distributed novels.”

io9’s motto is “We Come From the Future.” And misunderstand the present.

GBS and GSU: two cases going forward

Kevin Smith posted this [on March 23, 2011](#) at *Scholarly Communications @ Duke*. He focuses directly on what he thinks the future holds for the Google case. Excerpts:

Given the sweep of the rejection, and especially its finding that the “forwarding looking business model” is outside of the authority of the federal courts, this seems like a difficult decision to appeal. Nevertheless, I believe that it will be appealed, because I think the parties have very little choice. The other key part of Judge Chin’s decision, to me, is his strong suggestion that the settlement be converted to an opt-in agreement rather than an opt-out one. This would destroy its attraction to both sides, I believe, since it would exclude the ability to exploit orphan works. Without that huge financial opportunity, I don’t think settlement is worth it to either party.

Aside from reforming the settlement agreement in this way so that it could be approved by Judge Chin, the parties have two other options—continue the original litigation or appeal the rejection of the settlement as it stands. The first option seems unattractive to both parties at this point. Both would risk losing, of course, but more to the point, neither would have much to gain, at least not in comparison to the huge profit opportunity they think they have found in settlement. So I believe both sides will resist either returning to the original issue or reformulating the agreement in the way the Judge suggests and will instead appeal his decision, hoping to preserve that agreement more or less as it stands.

That’s not how things have turned out (at least so far), but it’s interesting to read Smith’s reasoning.

The Passive Virtues and GBS: Some Procedural Notes

Two more from James Grimmelman at *The Laboratorium*, [on March 23, 2011](#) and [March 26, 2011](#) respectively. The first discusses the relative brevity of Judge Chin’s rejection of GBS:

First, the structure of the opinion makes appeal more unpleasant. Chin didn’t put all his cards on the table. If the parties appeal his denial and win, one plausible outcome is that the case gets remanded back to him to try again—but he’s signaled that he’s likely to deny it again. That’s a long and protected litigation process, which can’t be encouraging to parties considering going the appeal route.

Second, by refusing to make new law on any issue except for Rule 23, he limited the uses to which his decision can be put as precedent...

The second harks back to the preliminary *approval* of GBS and two procedural consequences of that approval and Chin's eventual rejection:

First, the injunction against overlapping suits is now gone. If any authors or publishers don't like the plaintiffs' conduct of the suit and would prefer to go after Google directly, they're now free to again. It's quite possible that any such suits would rather quickly be transferred back to the Southern District of New York for combined processing, a bit like how the photographers' case was also assigned to Judge Chin. It's possible that some from the sow-Google-with-salt camp might choose this route, particularly if it seems that another settlement is in the offing.

Second, Paragraph 28 from the old order requires me to walk back my earlier assertions on Twitter that Judge Chin's order rejecting the settlement is not automatically appealable. His new order is also effectively a denial of class certification, which is immediately appealable. I'm not certain about this conclusion, in part because it comes after being certain about my earlier and opposite conclusion, and in part because the new order does not itself say that it denies class certification. But it does lead me to believe, in an "I would rather do something else on a sunny Saturday than research this further" kind of way, that if Google and the plaintiffs want to take an appeal, nothing stands in their way.

A little later (OK, a *year* later), Chin did indeed certify AG as a class; see earlier.

Legislative Alternatives to the Google Book Settlement

That's Pamela Samuelson's article in the *Columbia Journal of Law & the Arts*, deposited at SSRN [on April 25, 2011](#)—and James Grimmelmann's summary of it at *The Laboratorium* [on April 24, 2011](#).

The Samuelson article is 46 pages long. I will admit that I have not read it. Here's the abstract:

In the aftermath of Judge Chin's rejection of the proposed Google Book settlement, it is time to consider legislative alternatives. This article explores a number of component parts of a legislative package that might accomplish many of the good things that the proposed settlement promised without the downsides that would have attended judicial approval of it. It gives particular attention to

the idea of an extended collective licensing regime as a way to make out-of-print but in-copyright books more widely available to the public. But it also considers several other measures, such as one aimed at allowing orphan works to be made available and some new privileges that would allow digitization for preservation purposes and nonconsumptive research uses of a digital library of books from the collections of major research libraries.

James Grimmelmann is almost certainly the most significant source of ongoing coverage and commentary on GBS; he labels Pamela Samuelson (“Pam” to him) as “the most significant copyright scholar thinking about the Google Books Settlement.”

He considers Samuelson’s paper very important, saying “it deserves to be read alongside the discussions of a possible Digital Public Library of America.” Setting that aside for now, it’s almost certainly worth reading. I’ll quote Grimmelmann’s bullet list of key legislative elements raised by Samuelson:

- An expansion of the section 108 privileges for preservation, subject to appropriate safeguards such as security procedures. Digitization is an obvious and important component of preservation strategies; a well-crafted preservation privilege could help institutions like the HathiTrust use Google-scanned books to pass on our literary heritage. In a later section, Samuelson also argues for an expansion of library privileges in general. The Section 108 Study Group previously took a cut at this problem, but none of its (fairly modest) proposals have yet been acted on.
- A privilege to display snippets (subject to an opt-out) and to make what the settlement called “nonconsumptive uses” but Samuelson more accurately renames “nonexpressive uses.” (I would argue that both of these are or should be fair use already, but explicit recognition would provide a firmer legal footing.)
- “Congress should consider requiring Google to grant a license to other search engines to make nonexpressive uses of works in the GBS corpus.” Here, I wonder. I disagreed with the portions of Judge Chin’s opinion that could be read to suggest that Google’s initial behavior was necessarily reprehensible; Google engaged in activities that it reasonably thought were legal under copyright law. (I and others thought so, too.) Google’s competitors were not as tolerant of legal risks. This strikes me as a classic example of Learned Hand’s famous line from *United States v. Alcoa*, “The successful competitor, having been urged to compete, must not be turned upon when he wins.” In this case, if others would like to search the collected corpus of books, it seems reasonable to ask them to make their own scans. The real fix

here is to reform copyright law so that scanning for purposes of indexing is unambiguously legal—which is captured in Samuelson’s point about snippet display.

- Her proposal for what to do about orphan works is a clever compromise between the settlement and a full open-access regime:

Yet, Congress might consider adapting the GBS approach to orphan works to achieve a similar but better outcome. Congress could authorize the creation of an ECL for out-of-print books, as noted above; unclaimed funds from these books could be escrowed for a period of years; and after efforts to locate owners during those years failed, the works should be designated orphans and made available on an open access basis. If a book rights holder later came forward, he or she should be able to change the open access designation for such works.

- The mess over who owns electronic rights under decades of accumulated author-publisher contracts, in Samuelson’s view, is severe enough that it may justify Congressional action, perhaps along the lines of the settlement.
- The settlement’s programs for print-disabled readers were groundbreaking; similar provisions in copyright law in general would be a real breakthrough in meaningful access for a group that could most benefit from it.
- Privacy protections for readers are serious enough that they should be legislated.
- Finally, good-faith determinations that a work is in the public domain or was not commercially available should act as a shield from liability, provided that the entity stops treating it as such once the mistake is pointed out to it.

Speculating on the next GBS Settlement

In June 2011, it seemed to some observers that a GBS3 might be in the offing; thus, [this June 29, 2011 post](#) by Peter Brantley at *Shimenawa*—and it’s useful to remember that Brantley is deeply involved in these issues. He notes that Google wasn’t much interested in an opt-in settlement—one where rightsholders needed to explicitly agree to be part of Google Books. For that matter,

Arguably, not just Google would see diminished benefit from an all-parties opt-in regime for commercial uses. For many publishers, the existing Google Partners Program permits a degree of control over terms of access and revenue distribution that is unavailable through the settlement. At the cost of some bright-line clarity over author-publisher distributions associated with older contracts, publishers lose only the availability of an institutional subscription database (ISD); a

revenue model that is increasingly faulted for its coverage gaps as trade publishers pull out their more attractive titles, and academic publishers waver towards more open access principles under pressure from their host institutions and faculty authors. Additionally, academic catalog initiatives from Project Muse and JSTOR are likely to claim an ever-growing portion of university press backlists, and as trade backlist titles are digitized and enter markets at Amazon, Barnes & Noble, and Kobo, only smaller or niche publishers with fewer resources might benefit from settlement clauses. They are not the ones at the bargaining table.

Brantley also suggests AAP might be ready to “fold their cards” and go away—that continued litigation might seem like a bad idea.

This would leave the authors to negotiate with Google alone. It is not a far-fetched notion: the class action attorneys for the Authors Guild are operating under the premise that a settlement would fetch them their fair portion of an allocated \$45.5 million in attorney fees; there’s a clear financial incentive to see some kind of settlement emerge. But if it is to be authors only, what would an opt-in settlement look like?

He discusses some possibilities, and notes that an opt-in settlement would pretty much eliminate the Institutional Subscription because there wouldn’t be the huge database with full reading rights. He also wonders what happens to the Books Rights Registry in future scenarios, concluding that BRR is “the cobbler’s child that has no shoes (or perhaps only huaraches).” He concludes:

This discussion has attempted to illuminate one possible path forward; I present no assertion that this must be the road taken, and while directions such as this are being debated, the complex mix of factors and interests dictates hard against definitive analysis. Still, it is likely to be some form of reduced, hybrid model that emerges from the on-going discussions of the parties in the GBSS in the summer months ahead.

GBS: Settle or Litigate?

That’s Peter Brantley at *Shimenawa* again, this time [on July 22, 2011](#) after a second post-GBS2 status conference.

The parties indicated, nor surprisingly, that they needed yet more time, and that the slogging was tough-going. Judge Chin, in turn, indicated a bit of annoyance and suggested that they better move

on down a patch within a couple of months (by September 15, to be more precise).

At this point, Brantley doesn't see that Google would stand to gain much from a GBS3:

It seems to me that the only benefit Google obtains from a new settlement is clean hands over the past claims of infringement for digitization, but if the only operation they conduct is snippet-view, there is not necessarily a requirement for all-party approval. One could well argue from Google's perspective that they actually don't want to establish a precedent for asking permission for a broad class of activities that have been elsewhere held as Fair Use when they have been litigated. Furthermore, the barrier of final class certification resides primarily in the house of settlement; it need not be invoked if snippet display was decided on motion.

After another discussion, he arrives at a final paragraph that is both amusing and quite likely:

If the case should return to litigation in the absence of any settlement, even for claims of past infringement, there would be a number of potentially interesting consequences. One of those is that archives, museums, library associations, and the Internet Archive--the latter having been a particularly staunch opponent of the settlement--might actually wind up writing amicus briefs on behalf of Google in support of a favorable Fair Use finding. Far stranger things have happened in Silicon Valley.

Copyright politics make for oddly shifting alliances.

Divide and Conquer: Update on the Google Books Lawsuit

This article by George H. Pike appeared [in the February 2012](#) *Information Today*. Pike notes that the lawsuit (really lawsuits) is now seven years old—and recent events (and more noted in the previous section) “have kick-started the lawsuit from settlement talks back to the litigation process.” He suggests some directions that litigation might take, “ranging anywhere from a quick dismissal of the case to years of further litigation that could ultimately restructure U.S. and worldwide copyright law.”

He notes Google's “new divide and conquer strategy” aimed at removing the single massive lawsuit. That strategy has since failed. He also notes that nothing much seemed (or seems) to be happening on the AAP front.

Of course, the best solution would be for changes in the copyright law to reflect the technological changes and social benefit that the Google book database unquestionably provides. The orphan works

problem continues to loom; it inhibits not only Google but also any other organization that wants to digitize and make available any information that is copyrighted but does not have an identifiable owner. Millions of documents, photographs, works of music, and media items representing an extensive cultural and historical heritage exist in this netherworld, possessed by libraries and archives but limited to their dusty shelves.

Is it still likely that the lawsuit “could ultimately restructure U.S. and worldwide copyright law”? Would a finding in Google’s favor on the fair use issue have such an effect? I suppose we’ll find out over the next (few? many?) years. For now, life and the lawsuits both go on.

As does Google Books—although now it’s hidden under “More” on Google’s little black menu. As I write this, the phrase “Walt Crawford” yields “about 16,300 results” in Google Books on July 9, 2012, including most of my books (including self-published books) and, to be sure, some of Google’s special metadata sauce. *Library 2.0: A Cites & Insights Reader* shows up with a 2001 publication date (it was published in 2011)—but *it shows up* despite being only on Lulu and having sold no more than a dozen or so copies. For that matter, so does *DisContent: The Complete Collection*—and only five copies of that book exist, including the one on my bookshelf. That “about 16,300” turns into 349 as I page through the results. Why so many results? Sometimes there’s a character named “Walt Crawford” or one of the other semi-factual Walt Crawfords; in a surprising number of cases, I’m mentioned in a (usually library-related) book. I don’t seem to see too many snippets; in quite a few cases, there’s no way of knowing *why* the book’s there. (Jean Plaidy’s *The Sixth Wife*?)

In Closing

It’s been an interesting three years. This overview may be too long, but it’s as short as I felt I could make it while offering a range of representative viewpoints. I have no idea what the future will bring in the lawsuits, although I do believe another settlement is less likely—and that a settlement that covers so much more range than the cases itself is *really* unlikely.

Google Books should never have been touted as “the last library” or as a national library or the ultimate library or any of those things. Librarians should never have looked at GBS as an opportunity to stop housing physical collections while still being important. At best, GBS should have resulted in an interesting and potentially quite useful additional service. In any case, the settlement was doomed: It overreached fairness as a class-action settlement.

Your library isn’t going to be handed access to every book ever published. That probably wasn’t going to happen in any case. Life continues to be a little more complicated than that.

Masthead

Cites & Insights: Crawford at Large, Volume 12, Number 7, Whole # 151, ISSN 1534-0937, a journal of libraries, policy, technology and media, is written and produced irregularly by Walt Crawford.

Comments should be sent to waltcrawford@gmail.com. *Cites & Insights: Crawford at Large* is copyright © 2012 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/civ12i7on.pdf