



CPI Antitrust Chronicle

June 2011 (2)

After Google Book Search: Rebooting the Digital Library

Randal C. Picker

University of Chicago Law School

After Google Book Search: Rebooting the Digital Library

Randal C. Picker¹

I. INTRODUCTION

After more than a year of silence, on March 22, 2011, Judge Denny Chin rejected the proposed settlement in the Google book search (“GBS”) case. In the court’s view, the innovative settlement asked more than U.S. class-action rules could deliver and would, in Judge Chin’s words “simply go too far.” The settlement not only resolved possible liability issues for past acts by Google but would also have put in place an extensive forward-looking business arrangement. In the court’s view, Congress was the better forum for establishing the new regime set forth in the settlement agreement. The court did suggest a path forward but one that would undermine many of the potential benefits of GBS.

The rejection of the settlement (known formally as the Amended Settlement Agreement or “ASA”) means that we are at a point of rebooting how we design our digital library future. There were many criticisms of GBS and the settlement but perhaps chief among those was the risk that approval of the settlement would have locked in a single approach to digital libraries. Google would have received unique access to the so-called orphan works and that would have provided it what may have been a decisive advantage against digital library competitors, both private and public. As we move forward on the orphan works, we need to do so with two principles in mind. First, we need to enable broad competing uses of the orphan works while, to the greatest extent possible, respecting the rights of the orphan works holders. Second, we should not repeat the mistake of the GBS settlement by somehow tilting the table in favor of digital library monopoly, either public or private.

We should want to foster a rich digital library ecosystem. GBS makes clear that we can have large-scale private digital libraries. That is an important development and one that we should seek to enable. If we create use rights for copyrighted works for digital libraries, we should be sure to make those privileges available to both public digital libraries and private digital libraries such as GBS and its successors. Our existing statutory safe harbors for libraries favor noncommercial libraries and archives. The emergence of GBS suggests that that is too narrow a conception of what libraries can be in the digital age and we need a statutory scheme that supports that.

II. GOOGLE BOOK SEARCH: THE VISION AND THE REALITY

We should return to the beginning with the promise of Google Book search: Google would digitize the world’s books and make them available online. The digital library of Alexandria. The vision was clear, the reality quite a bit different, but I will focus on the legal

¹ Paul H. and Theo Leffmann Professor of Commercial Law; Senior Fellow, the Computation Institute of the University of Chicago and Argonne National Laboratory.

issues here. Anyone who has stood in front of a Xerox machine and copied an entire book—that would be me, I confess, though not in this millennium—gets the idea that copying an entire book looks like a copyright violation. Multiply that by millions and then add injury to injury by making parts of the books available online. As I will discuss later, the copyright issues are more interesting than this description suggests but Google had to expect that it would get sued and the Author's Guild did just that through a class-action lawsuit.

It was the proposed settlement of that lawsuit that was before Judge Chin. The ASA was complex though not unduly so given what it was trying to accomplish. If the Author's Guild was right, Google had engaged in widespread copyright infringement and copyright holders were entitled to damages and possibly statutory penalties that could easily have run into the millions. A typical class-action settlement would need to resolve the question of liability for these past acts. But you can't build a business or create a great resource just resolving past liability. The ASA would have put in place an ongoing arrangement that would have enabled Google to use copyrighted works as it was doing in GBS and to generate revenues that would be paid to copyright holders.

The settlement was organized as an opt-out settlement and this was critical to the vision. This case was always about default settings. Google could accomplish large chunks of what it sought through contract and it has done so. Active authors or publishers who believed that the generic settlement didn't work for them would opt out and cut a separate deal. These rights holders couldn't be forced to be in the settlement and always had the possibility of a separate contract as an available alternative.

Of course, for works in the public domain, Google did not need the consent of copyright holders. Instead, Google needed to figure out a means of accessing those works and digitizing them and Google did exactly that often through agreements with university and public libraries. More contracts but with different parties.

But Google could not rely on contracts to use the orphan works, that is, the works without readily identifiable copyright holders. The genius of the settlement was precisely the way in which it surmounted the consent requirement associated with many uses of a copyrighted work. The opt-out class action offered the chance to flip the default position so that orphan work holders had to opt out affirmatively of the settlement and, through that, Google offered a path for its use of the orphan works.

III. CONGRESS, NOT THE COURTS

This was the settlement that the court faced. The opinion starts with background on the case and the settlement and then ticks off seven areas of objection to the settlement: (1) adequacy of class notice; (2) adequacy of class representation; (3) scope of relief under Federal Rule of Civil Procedure 23; (4) copyright; (5) antitrust; (6) privacy; and (7) international. Three of those relate to class action law, three to substantive areas of U.S. law, plus we layer on top of that multi-jurisdictional concerns. The underlying standard for the settlement of class actions looks to whether the settlement is fair, adequate, and reasonable. That isn't particularly precise, but we may not be able to do better and instead need to rely on the experience of judges in confronting these sorts of situations.

The opinion quickly rejected the challenge regarding inadequate notice. Notice matters, of course, but it isn't that interesting and there was no real suggestion that something nefarious was afoot with regard to the notice. The class representation inquiry is much more interesting.

The key to this case was the possibility that Google might acquire a license to use the orphan works without having to get the consent of the copyright holders of those works. This wasn't about an unwillingness to get that consent, but rather the sheer inability to negotiate with unknown persons. There can be reasonable disagreements about the extent of the search that should be undertaken before a work is treated as having orphan status, but wherever we draw that line, we clearly will have some orphan works.

That generates an obvious question about representation; namely, can the active copyright holders bringing the lawsuit fairly represent the absent orphan rights holders? Framed that way, we have split the groups neatly but that, of course, doesn't mean that that categorization is legally meaningful. Google and the Authors Guild could have moved to have a guardian *ad litem* appointed in the case to represent the orphan rights holders. We see exactly that sort of appointment in other complex cases, such as asbestos bankruptcies, where current tort victims may have different interests than future tort victims. Appointing an independent representative for the orphan works would undoubtedly have made the negotiations more complicated, but that is the precise point: it is easy to get one side to agree when they aren't actually represented.

But there is a second issue on representation and it runs in conjunction with the first. An opt-out class action is exactly that: members of the class can reject the settlement and seek a separate deal with Google. That is run-of-the-mill stuff in class actions but what we should fear here is that substantial numbers of copyright holders opt out and cut a better separate deal. There is more reason to do that here than in normal class actions precisely because the real project of the settlement is to build a business and not just settle lawsuits relating to past acts. Google will want to have in place arrangements for new books as those come out. However, those rights won't come out of the settlement, but will need to be established through separate contracts. In the extreme version of the case, large publishers opt out of the settlement and the only rights holders left in are smaller publishers and orphan rights holders. The opinion didn't address these issues in great detail but instead regarded these issues as "substantial" and "troubling."

Instead, the opinion focused on whether the class action rules allowed the going-forward arrangement that GBS represented. Judge Chin understood these to be a question of comparative institutional competence: "The questions of who should be entrusted with guardianship over orphan books, under what terms, and with what safeguards are matters more appropriately decided by Congress than through an agreement among private, self-interested parties." There is much more in the opinion—we are now on page 23 of a 48-page opinion—but that is the heart of the analysis. Congress, not courts, should resolve the problem of orphan works.

There is more on class actions, namely whether the settlement would be within the scope of the case as originally framed in the pleadings. This is another flavor of the court's concern over using class-action law to build a business. The pleadings originally had been about Google's indexing of works to respond to search requests and the use of those works to display responses to those requests, a far cry from a settlement which contemplated selling access to entire books.

Turning to the objections based on U.S. substantive law, Judge Chin's opinion makes two key points: First, that many authors had objected to the settlement and, second, that forcing the copyright holders to opt out was inconsistent with the fundamental organization of U.S. copyright law. These points are tied together. Rights holders who didn't object to the settlement

could be split into two groups: non-orphan rights holders who found the settlement acceptable and orphan rights holders. As Judge Chin noted, there is no obvious reason to think that orphan rights holders would have had a different perspective on the settlement. That would suggest that rather than treat all orphan holders as accepting the settlement, we should have treated them as rejecting the settlement in precisely the same fraction that the settlement was rejected by active holders. The opt-out class action instead treats all orphan holders as accepting the settlement, though the settlement did preserve a later right to opt out.

The antitrust analysis in the opinion was short—three double-spaced pages—and made two points. First, the settlement “would give Google a de facto monopoly over unclaimed works.” The Department of Justice had pressed this point in its filing and while it is hard to disagree with the response to that—1 is more than 0 so we are better off with the deal than without it—that response doesn’t really confront the question of when, if ever, is it appropriate for the government to create a monopoly license? Judge Chin though also focused on what the ASA meant for a second, adjacent market: “The ASA would arguably give Google control over the search market.” The opinion is full of hedges and doesn’t do anything like the kind of full analysis we would expect in an antitrust case.

And the court dealt with privacy issues in an even more truncated fashion. “The privacy concerns are real,” but the court didn’t think that they were sufficient to reject the settlement. Instead, the court thought that undefined adjustments could be made to protect privacy while still allowing Google to engage in “marketing efforts.” The opinion cited no relevant privacy laws so we actually learn very little from the opinion on the privacy issues. Finally, the court offered an extended description of objections by foreign rights holders but little by way of analysis. In the court’s view, the foreign objections offered yet another reason why Congress was the preferred forum for the resolution of these issues.

With all of that said, Judge Chin rejected the settlement. He did suggest that converting the settlement from opt-out to opt-in would solve many of the objections raised. Of course, as I am sure that he recognized, an opt-in settlement would be little different than what be accomplished through contracts. Most importantly, this would leave the orphan works sitting on the sidelines, unless Google was willing to continue to use them in reliance on its fair use claims under copyright.

IV. THE DIGITAL LIBRARY ECOSYSTEM

Where does that put us? The official settlement website, which can be found at www.googlebooksettlement.com, notes “the parties are considering their next steps.” At a status hearing on June 1, 2011 the parties asked for more time to do just that and were given until July 19, 2011. We need to separate possible liability for past acts from going forward operations. An opt-out class action for past liability would be quite conventional. As to going forward, through its books partner program, Google has put into place extensive contracts that will enable whatever uses of those works copyright holders permit. As to the orphan works, absent legislation—more on that in a second—Google faces some choices.

Were I Google, I would want to distinguish between the use of works to improve its search engine from the presentation of chunks of the work to the public (so-called “snippet use.”) The search-engine use—in the language of the case “non-consumptive research use”—may very well stand on a different footing than snippet use and we should not just assume that the fair use analysis will apply equally to all possible uses of the works in question. It is one thing for a human

being to read the works of, say, Ernest Hemingway and quite another thing to have a computer process the text to understand word usage. If a key point of the project for Google was to improve its core search engine to better compete in the search engine market, Google may be able to get much out of what it wanted from the orphan works pursuant to fair use without ever displaying those works to the public.

The at least temporary disruption of GBS has re-energized ideas for alternatives to it. Why not move the public library online? If we have had physical libraries financed through general tax dollars and free to the public, why can't we do the same thing with online digital libraries? Some are calling this the Digital Public Library of America ("DPLA").² While GBS has been free so far, the settlement contemplated that Google would move to charging for broad access to GBS, though the settlement did contemplate one free terminal at each physical public library. Indeed, one of the criticisms of the GBS settlement was precisely that Google intended to charge what many feared would be a high price for this access. Perhaps far better to create a genuinely free online public library.

Take stock on where we are right now on that project. The American Library Association issues an annual report entitled "The State of America's Libraries." In its 2011 report, the ALA noted that 94 percent of all academic libraries are offering some eBooks, as are 72 percent of public libraries.³ Books are being circulated as downloads or preloaded on reading devices. Actual eBook circulation figures are still small in number, but growing. The Chicago Public Library reported eBook circulation of 17,000 in 2009 and more than 36,000 in 2010. These eBooks are direct substitutes for physical books and haven't required a change in the way in which libraries purchase books. The mechanics on lending and check out are a little different, but the core idea is straightforward: The library buys a certain number of digital copies and a patron can check out the book if one of those copies is available on the virtual shelf.

But GBS is really a different creature, as presumably would be a significant public competitor. This is a searchable database of scanned books, not just a library of digital books with searchable metadata. The GBS settlement called for a split of the revenues that the project would create—roughly 37 percent to Google and 63 percent to rights holders. A free online public library version of a book database wouldn't generate revenues of this sort. It seems unlikely that copyright holders would be satisfied selling one never-checked-out, searchable copy of a digital book to the DPLA for \$14.99 or whatever eBooks sell for these days.

As that should suggest, the contracting process for in-copyright works with active rights holders won't be simple. Copyright holders will be looking for revenue streams and will have the full right to prefer revenue-generating services like GBS to a free online public library. Many electronic databases are sold today with lump-sum payments, but that wasn't the model of GBS and we should be skeptical that copyright holders will want one-time, lump-sum access fees. Plus, if they were willing to do so, we might be nervous that the government might be able to play favorites through its purchase prices.

Instead, I suspect that we may see digital public libraries track the deal in GBS by paying on a usage basis. The 37/63 revenue split in the ASA is a usage deal, just one tied to revenues. We could imagine public libraries that charged for the use of books—financed through user fees

² See Robert Darnton, *Google's Loss: The Public's Gain*, N.Y. REV. BOOKS, (April 28, 2011).

³ THE AMERICAN LIBRARY ASSOCIATION, *THE STATE OF AMERICA'S LIBRARIES*, p. 36 (2011).

(including the possibility of price discrimination) rather than through general tax revenues—but that would be controversial and, in any event, the public library would probably be operated on a non-profit basis. Revenues aren't likely to be the basis for usage fees paid by public libraries to copyright holders. Instead, we might expect deals that tracked use, though of course that would require us to quantify use and to deal with the equivalent of click fraud. Having usage metrics and standardized fees will help to sidestep the favoritism problem identified above.

We can now see the tradeoffs we face regarding private and public online libraries. Private libraries are just that and are likely to limit access to those willing and able to pay, though even the GBS settlement contemplated some free public access to the books database. Public libraries are likely to facilitate broad access, an important democratic value. The problem with a new online public library isn't with its users but is, instead, possible issues in the acquisition of new works. Copyright holders will be nervous that ease of use of a digital public library will mean that consumers will substitute out of buying books. Public library eBooks are at an early stage, but as numbers have started to grow, publishers are adjusting how they approach eBook sales to libraries. In a move that generated widespread discussion, in February, 2011, HarperCollins announced that going forward its library eBooks would expire after 26 checkouts.⁴ Physical books degrade but digital books have a much longer natural lifespan. Publishers and libraries have opposing interests here. Publishers will want to preserve revenues—or at least net profits—through the transition to digital while libraries will look for cost savings.

We are clearly at an early stage in our transition to digital libraries. All of the above suggests that we need to expand our conception of what a library is and that we should not tilt the legal tables in favor of public or private libraries. This issue matters most obviously as we circle back to the problem of orphan works. If Judge Chin is right that this is a problem for Congress, then we need to figure out what that legislation should look like. Only the government can create a license for those works and I am hard-pressed, as a matter of first principles, to understand why that license should be limited. That means that a license should not run in favor of one party, nor should it be limited, as suggested by Robert Darnton, to entities that wish to make noncommercial uses of those works.⁵ New orphan-works legislation should enable broad competing uses of the orphan works, by both commercial entities and non-profits.

To the extent that we create other statutory helps for digital libraries, we need to ensure that we enable private efforts as well. The current copyright statute draws sharp lines here—mainly in Section 108—in carving out special exemptions for noncommercial libraries and archives. The noncommercial limit undoubtedly taps into a sense that we shouldn't do special favors for those seeking to profit from copyrighted works, but we need to step carefully here. I confess to skepticism about creating special copyright exemptions to subsidize noncommercial libraries. If we believe in the subsidy, then distribute the burden of it generally and don't just target copyright holders. It would be easier to run public libraries if they received free paper and pens but we don't require Office Depot to ship stuff to the libraries for free. But even if you buy the notion of a special in-kind copyright subsidy for public libraries, that isn't to say that we shouldn't also create safe harbors for private, for-profit libraries.

⁴ See Calvin Read, *Librarian Unhappiness Over New Harper e-Book Lending Policy Grows*, PUBLISHERS WEEKLY, (March 2, 2011).

⁵ Robert Darnton, *A Digital Library Better than Google's*, N.Y. TIMES, (March 23, 2011).

What does that mean operationally? There is a great deal to work out regarding the mechanics of orphan works legislation. We have had draft bills in the past and a comprehensive report by the Register of Copyrights on orphan works and now a post-GBS literature is emerging.⁶ The animating principle of such legislation should be to try to replicate what we think orphan rights holders would do were they actually present. Doing so would exhibit the greatest fidelity to the existing copyright system. Orphan-works legislation shouldn't be seen as an opportunity for giving orphan holders weaker rights merely because they aren't present and are unrepresented.

For example, copyright holders don't typically just give books for free to public libraries. If orphan works are included in an online public library, the government should escrow payments for those works similar to the payments it makes for comparable present rights holders. Those escrow payments might escheat back to the government under normal rules should orphan holders never come forward. A weaker version of this idea would be to exempt public libraries from these payment obligations for orphan works and to only make for-profit libraries pay. Again, I don't see the basis for nonpayment by public libraries, but to the extent that there is support for that idea, we shouldn't conclude we need to exclude for-profit libraries from access to the orphan works. Much better to run a two-tier system: free access to the orphans for public libraries; fee access for private libraries.

Obviously, a digital library and database needs digital books. There are a number of competing U.S.-based scanning efforts, plus other initiatives around the world. One leading example is the HathiTrust, which is a consortium of research libraries and currently has roughly 8.7 million digitized volumes.⁷ Its digital library includes both public domain works—27 percent of the total volumes—and in-copyright works, and it provides full-text search across its entire library. For in-copyright works, it returns only page numbers indicating where the search term has been found but, unlike GBS, it does not show the search term in context in the work. For public domain works, it shows search results in context and offers downloads. It is important to remember that a good chunk of the scans in the HathiTrust are from Google's scanning efforts, so we shouldn't think of this yet as large-scale alternative digitization.

We should want competing approaches to privacy, scanning, metadata, and search. There will be a temptation to leverage the scans that Google has already done.⁸ As I have indicated before, I don't see the basis for that. Antitrust proper imposes few mandatory-dealing obligations on a single firm. Copyright does create in-rem remedies, so a great deal turns on how we see the original copyright case against Google. Finally, the government could turn to eminent domain and pay for access, though doing that raises interesting price and public use issues.

V. CONCLUSION

This is a fascinating time in libraries. The switch from physical libraries to digital libraries means that we are at an interesting stage of institutional design. GBS is likely to move forward in one form or another. The most limited version would include the entire public domain plus whatever works Google can negotiate access to through contract. Orphan works legislation could

⁶ The U.S. Copyright Office's orphan works website includes links to reports, draft legislation, and testimony, *see* www.copyright.gov/orphan/. As to post-GBS legislative suggestions, *see* Pamela Samuelson, *Legislative Alternatives to the Google Book Settlement* (online at <http://ssrn.com/abstract=1818126>).

⁷ *See*, www.hathitrust.org.

⁸ *See* both Darnton, *supra* note 5, and Samuelson, *supra* note 6.

greatly add to what Google and other libraries, public and private, can offer. We are likely to see public and nonprofit efforts, both here and abroad. Orphan works legislation should respect the rights of copyright holders to the greatest extent possible while enabling use of those works.

The great problem with the amended settlement agreement negotiated between Google and the Authors Guild was precisely in the way that it seemed to tilt the tables powerfully in favor of one, and only one, model of the new digital library. We should want this ecosystem to be rich and teeming. Both public and private efforts are likely to have distinct advantages and disadvantages and we should be sure that the government doesn't resolve the institutional design question through casual fiat. Approving the ASA might have done just, but we have now sidestepped that. Enacting orphan works legislation that somehow only allows noncommercial libraries to use those works would be to commit the same mistake, just in a different form. More broadly, legislation enabling new digital libraries should foster digital libraries generally and should operate from a posture of neutrality as to whether those libraries are public or private or non-profit or for-profit.