

HOW TO FIX THE GOOGLE BOOK SEARCH SETTLEMENT

By James Grimmelmann

The proposed settlement in the Google Book Search case should be approved with strings attached.¹ The project will be immensely good for society, and the proposed deal is a fair one for Google, for authors, and for publishers. The public interest demands, however, that the settlement be modified first. It creates two new entities—the Books Rights Registry Leviathan and the Google Book Search Behemoth—with dangerously concentrated power over the publishing industry. Left unchecked, they could trample on consumers in any number of ways. We the public have a right to demand that those entities be subject to healthy, pro-competitive oversight, and so we should.

After laying out the basics of the proposed settlement agreement and explaining how it will benefit both the parties to the lawsuit and society more broadly, I'll work through the implications of five overriding principles:

1. **The Registry poses an antitrust threat.** The Book Rights Registry will be a new collecting society representing the interests of authors and publishers. This kind of collective action poses an obvious antitrust risk: What if the Registry becomes the instrument of a cartel to fix the price of books?

To keep the Registry from overreaching, we'll need more checks than the settlement currently contains.

2. **Google poses an antitrust threat.** By virtue of the settlement, Google will have the book search market and the download market for orphan works largely to itself. The structure of a class action settlement, in itself, is a highly effective barrier to entry. We'll need similar checks on Google's actions to keep it from acting anticompetitively in book markets.
3. **Consumers need protection.** The settlement as it stands is often good at ensuring minimum standards to benefit libraries, institutions, and consumers. But in some areas, such as price discrimination and privacy, the settlement leaves the door open for Google to behave oppressively. We'll need to close that door.
4. **Public goods should be widely available.** Just by providing the search and download services and

Continued on page 11

James Grimmelmann is Associate Professor of Law, New York Law School. An earlier version of this article was originally published on the *Laboratorium* weblog at http://laboratorium.net/archive/2008/11/08/principles_and_recommendations_for_the_google_book, on November 8, 2008. Aislinn Black, Joseph Gratz, Joris van Hoboken, Fred von Lohmann, Alexander MacGillivray, Dominic Mauro, and Derek Slater gave helpful comments. This article is available for reuse under the Creative Commons Attribution 3.0 United States license, <http://creativecommons.org/licenses/by/3.0/us/>.

IN THIS ISSUE

HOW TO FIX THE GOOGLE BOOK SEARCH SETTLEMENT	1
<i>By James Grimmelmann</i>	
LAW IN VIRTUAL WORLDS.....	3
<i>By Rónán Kennedy</i>	
INTERNET LAW IN THE COURTS	21
<i>By Evan Brown and Brian Beckham</i>	

How to Fix the Google Book Search Settlement Continued from page 1

by processing payments, Google and the Registry will assemble some immensely useful databases about book copyright information. These databases are classic public goods, and neither Google nor the Registry will need exclusive rights over them as an incentive. We'll need to make sure that those databases are made available to the public.

5. **Transparency and accountability matter.** If Google becomes a chokepoint for getting books to the public, there's a risk that it could secretly censor. Google, of course, doesn't want to distribute books that it strongly disagrees with (or fears legal liability from). These two goals can be reconciled. We'll need to make sure that when Google chooses not to make books available, it leaves open suitable alternative channels for getting them to the public.

These five principles give rise to a number of specific recommendations. Each proposed change is incremental and fully consistent with the settlement's overall goals and design.

THE SETTLEMENT

First, a little history. Google Book Search is actually two related programs. The Partner Program works with books that it receives from publishers (either physically or as PDFs). Meanwhile, the Library Project has been scanning books en masse from the collections of university and civic libraries. Either way, Google folds the scans into its gigantic Book Search index. Enter a search term, and Google will show you a list of books containing the term.

What happens when you click on a result depends on where the book came from. For public-domain books from the Library Project, you go directly to a page where you can browse the book online or download a PDF copy. For books from the Partner Program, Google lets the publisher decide how much users can see. Generous publishers enable extensive previewing of sample chapters; stingy publishers allow you to see only a snippet of a few lines containing the search term; paranoid publishers prohibit even that.

This leaves open a large middle ground: in-copyright books from the Library Project. Most of these books are out of print; many of them are "orphan works" for which no copyright owner can be found. Google's policy has been to make these books searchable and available in snippets. That may not seem like much, but it's still more

than many authors and publishers would like; they see Google as a trespasser willing to drive through the yards of people who aren't at home.

The proposed settlement is complex, but its central deal is simple. Looking backward, Google will be released from liability for its scanning, searching, and displaying² in exchange for a set of one-time payments totaling about \$125 million.³ Looking forward, it will be allowed to continue scanning and displaying books⁴ in exchange for 63 percent of its net revenues from the advertising that it shows on search results and book display pages.⁵ This deal is most significant for out-of-print and orphan works: It effectively lets Google go ahead with making substantial uses of these works, while holding a share of the revenues in trust for the copyright owners.

Of course, many of the copyright owners who'll be paid as part of the settlement don't currently have deals with Google. A new Book Rights Registry will be in charge of mediating all these relationships.⁶ It will take payments from Google and pass them along to the appropriate authors and publishers. In the other direction, it will communicate to Google copyright owners' requests for the levels of access allowed to their books.

The settlement also authorizes some exciting new programs. An Institutional Subscription system will allow Google to sell all-you-can-eat digital access to the entire catalog of scanned books to companies, colleges, and potentially even individuals.⁷ A Public Access Service will provide a restricted version of the Institutional Subscription to colleges and public libraries for free.⁸ Meanwhile, a Consumer Purchase system will allow Google to sell electronic access to complete individual books.⁹ Copyright owners can set their sale price or delegate that decision to Google's algorithmic whizzes.¹⁰ The revenues from these new programs, along with other new revenue models worked out by Google and the Registry, will also flow to copyright owners in the same 63/37 split.¹¹ A section on "Non-Consumptive Research" also allows researchers to run gigantic automated statistical studies on the entire corpus of scanned books.¹² Copyright owners won't be paid for these uses on the theory that they don't involve people "reading portions of a Book to understand [its] intellectual content."¹³

Meanwhile, the libraries that have been supplying books for scanning aren't actually parties to the lawsuit. The settlement, however, contains provisions that enable them to sign agreements with the Registry that effectively release them from liability.¹⁴ Libraries that get back digital copies of their books from Google must agree to keep those copies secure and limit access;¹⁵ libraries that don't have few other obligations.¹⁶

THE SETTLEMENT SHOULD BE APPROVED

My starting point is that the settlement is a good thing. Everyone is better off than they would be in a world without Google Book Search:

- Google makes piles of money from selling ads, subscriptions, and e-books.
- Authors and publishers get the lion's share of that money.
- The publishing industry benefits from the incidental cleanup of copyright records.
- Universities, schools, companies, and other institutions can subscribe to Google's convenient fire hose of books.
- Public and college libraries get free access to the fire hose.
- The libraries participating in scanning books get digital copies of the scanned books from their collections.
- Individual readers get an increasingly comprehensive book search engine.
- Readers also get free PDF access to millions upon millions of public-domain books and two paid-but-convenient digital sources for in-copyright books.
- The public as a whole gets a substantial leg up on solving the orphan-works problem.
- Researchers running automated studies can advance human knowledge on algorithms, natural language, the history of publishing, and other topics.

These are serious benefits, and the settlement is a universal win compared with the *status quo*. Still, some commentators dislike the settlement because they think we'd be even better off if the case proceeded to trial and judgment. They'd been hoping that Google would have established that its scanning and searching features were fair uses. If Google had prevailed on the fair use issue, it would have opened the book search business to anyone, free from legal taint. What's more, it would have given us a powerful, portable fair use principle that could do a lot of other good in this digital age. While a definitive finding of fair use would have been better than the settlement, that's not the choice on the table. There are three good reasons that we should settle for the settlement.

First, there's no guarantee that Google would have won on the fair use issue. A finding against fair use, especially given the possibility that the Supreme Court would have granted *certiorari* in such a socially significant case, could have done immense damage to other large-scale projects making beneficial but incidental uses of a great many copyrighted works. The settlement is so

comprehensive, however, that it gives us perhaps 80 or 90 percent of the actual uses of books that a positive fair use finding would have enabled. Those who favor a judicial showdown on the fair use issue would do better to join that battle in some other case, one in which the equities tip more clearly in favor of fair use.

Second, even if scholars and observers would like a fair use fight, it's not our call to make. It's Google's. Google was the defendant; it earned that dubious privilege by actually scanning and searching books. Having stepped up to the plate to risk a lawsuit, and having been beamed with one, Google now has the right to choose whether to settle that suit. Google's choice to settle takes away no legal rights from anyone else; one else loses the fair use argument because Google didn't chance it.

I've seen people raise the argument that Google's capitulation means that there's now a functioning licensing market for searching books. If true, this fact would undercut fair use claims by Google's competitors, since it would imply that scanning and searching without payment take away revenue that copyright holders could have realized. The critical hole in this argument is that this isn't a market that one can effectively negotiate in without the device of the class-action lawsuit. Even after the settlement, there will still be far too many potential plaintiffs for any competitor to be able to cleanly license all the rights that it needs to start large-scale scanning. It's also significant that Google won't pay royalties for scanning or searching under the settlement; the paid uses only start once a user clicks through from a search results page to a page with content from a specific book.¹⁷ If the settlement creates a distinct market for scanning and searching, it's a market in which these uses are worthless to copyright owners.

Third, given that Google and the copyright owners are the parties here, there's no way to force the court to consider the fair use issue. The adversarial legal process allows parties to present their cases and controversies to a judge for resolution. It doesn't generally allow outsiders to compel the parties to litigate issues not of their choosing. This fact means that, to the extent that the rest of us want the court to modify the settlement or even to think about particular issues, we need to find a hook to put them properly before the court. There are some such hooks, but the fair use question isn't one of them. There is no convincing way to categorize that specific copyright issue as part of the general public interest that the court is directed to take into account in approving a settlement.

Thus, I start from a baseline of believing that the settlement should be approved. It makes all of us better off, and it is, in an intuitive and meaningful sense, quite fair to all involved. I have concerns and critiques, but I see

them as patches to make the settlement better, not as do-or-die clauses that the settlement must absolutely contain. My suggestions flow from the principles identified earlier.

PRINCIPLE 1: THE REGISTRY POSES AN ANTITRUST THREAT

The basic deal embodied in the settlement—authors and publishers who don't opt out will receive payments out of Google's revenues—will create a huge administrative workload. Someone needs to maintain a database of who owns which copyrights, mediate in ownership disputes, process payments, audit Google, and so on. Although collecting societies currently do many of these jobs for other kinds of works, such as ASCAP for musical works and SoundExchange for sound recordings, the United States doesn't currently have a suitable similar organization for books. Nor can these tasks be trusted to Google, which would have an obvious conflict of interest if it were charged with tracking down AWOL authors or being its own auditor. Thus, the settlement agreement establishes a new collecting society, the Book Rights Registry.¹⁸

The Registry doesn't just have ministerial tasks, though. It also has substantial authority to negotiate on behalf of authors and publishers. It has approval power over the security standards that Google and the various libraries must create and live by.¹⁹ It has broad discretion to work out an equitable formula for dividing revenues among publishers and authors.²⁰ It can even negotiate the terms of new revenue models (*e.g.*, print-on-demand, PDF downloads, and coursepacks).²¹

If your antitrust sensors aren't pinging wildly at this point, please make sure that they're properly calibrated. The Registry is a centralized entity with the authority to negotiate on behalf of all registered book copyright owners. As such, it walks and quacks like a cartel. There's a reason that ASCAP and BMI, which play similar roles for musical compositions, live under antitrust consent decrees; otherwise, they could make or break radio stations and musicians and drive up the effective price of music. Were all authors and publishers (*i.e.*, the plaintiff class, more or less) to sign a piece of paper giving the Registry these powers, it could barely lift a finger without violating the Sherman Act. That fact doesn't change just because the agreement is the result of a class action settlement rather than a meeting in the world's largest smoke-filled room.

The settlement recognizes this danger, and accordingly puts some very important limits on the Registry. It's specifically prohibited from representing any subgroup of copyright owners; it has to act in all their interest.²² (This restriction keeps it from being used by one group of authors, say, to suppress the market for another group's

books.) Similarly, its board is equally divided between authors and publishers, with any action requiring a majority.²³

Further, the settlement ensures that the licenses that copyright owners grant to participate are non-exclusive; they can strike side deals with anyone they like.²⁴ They're guaranteed more than two years to decide whether to remove their books from the program entirely.²⁵ Even after that, they can opt out of almost everything beyond basic searchability, in which Google will tell users what page of a book contains a search term, but won't show them any of the actual text around the term's appearance on that page.²⁶ Even within the Consumer Purchase system, individual copyright owners are always free to set the price for e-books sold through Google.

These facts keep the Registry from acting like a classic price-fixing cartel; individual publishers can easily defect and charge less (or more). It's true that the Registry could facilitate coordination, but that's not a big issue in the publishing industry. Publishers can already see what each others' suggested prices are by looking on Amazon.

The Registry, however, will be an anticompetitive threat for the same reason that this lawsuit will transform publishing: the settlement's class-binding effect. If Google would like to negotiate, say, an encryption standard and DRM terms for book downloads, without the Registry, it needs to negotiate one-on-one with authors and publishers. But the Registry is authorized to negotiate on their behalf, all of their behalf. It could agree with Google on a privacy-intrusive DRM standard that fed back usage information into a database used to do industry-wide price-fixing in the guise of price discrimination. Other examples abound; the Registry's centralized negotiating role permits various anti-competitive practices to be coordinated and laundered. What should we do about these risks?

- **Antitrust consent decree.** The Registry should be under ongoing antitrust supervision from the day of its birth. The Department of Justice (DoJ) should require that it negotiate and sign an antitrust consent decree. That decree would enumerate various forbidden anticompetitive practices, including those called out in the settlement, ones pertaining to hidden price-fixing, and whatever else the experts in the Antitrust Division think necessary to add. In addition, the DoJ should be given the authority to review all contracts entered into by the Registry and reject any with anticompetitive effect.
- **Nondiscrimination among copyright owners.** For understandable but regrettable jurisdictional reasons, the plaintiff class doesn't include owners of unregistered copyrights, nor does it include future

authors. Only currently registered copyright owners are part of the lawsuit, which means that only they can be part of the settlement.²⁷ The most immediate danger is that the Registry might adopt policies that operate to the benefit of past authors and against future authors (policies with narrow views of fair use and broad views of a derivative works right come to mind). There's an easy way out, which is that the Registry must be explicitly required to represent any copyright owners who agree to its standard deal and explicitly forbidden from offering non-plaintiff and future copyright owners any materially different deal. This pair of rules guarantees to anyone who comes along in the future effectively the same opt-out right that the settlement class enjoys, while offering them exactly the same terms if they choose not to opt out. This way, the Registry will truly be a fair, impartial representative of all authors and publishers.

- **Library and reader representation at the registry.** Beyond that, the Registry's structural protections should be supplemented to include voices from outside the publishing industry. It's acceptable and understandable for the Registry's charter to require that at least one author representative and at least one publisher representative consent to any action that it takes, but that veto principle doesn't require that they be the only members of its board. It should also contain members representing libraries and the reading public. In addition to objecting if the Registry takes anti-competitive, anti-reader actions, these additional members would be able to monitor the Registry's actions, bringing important transparency to this new quarter-ton gorilla of the book industry.

PRINCIPLE 2: GOOGLE POSES AN ANTITRUST THREAT

The Registry isn't the only entity that'll have market power as a result of the settlement. Google will too. It'll become the only game in town for scanning and searching books on anything resembling this scale. Yes, it was the only game in town on this scale before, but that was when there was a legal threat hanging over it. Now Google will have the legal okay to go full-steam ahead, with some exceedingly tasty markets all to itself.

The immediate rejoinder from Google, of course, is that these markets aren't closed to entry. Microsoft used to have a book-scanning program; it could have one again. There's nothing in the settlement to prevent anyone involved from doing side deals with others. Authors could license Yahoo! to scan, index, and sell. The Registry could split the take. Libraries could subscribe to Yahoo!'s

version of the fire hose. The settlement just sets up a series of deals with Google; it leaves everything else open for all comers.

The problem with this argument is that you can't actually just go out and do what Google has done. One of Google's fair use arguments would have been the insane transaction costs of trying to negotiate with every possible copyright claimant, particularly for out-of-print and orphaned books. The settlement gives Google a clean release from the transaction-cost madness. All those pesky claims from authors who can't be found or won't play ball just go away.

Consider the unappetizing options facing a would-be competitor like Yahoo! If it goes ahead and starts doing large-scale scanning, it'll get sued just the way Google did, but there's no guarantee that the plaintiffs there would feel any interest in settling on terms comparable to the ones that Google got. Indeed, as long as there were potential plaintiffs out there, Yahoo! couldn't feel safe, even if it had struck agreements with 99 percent of them. The others could still pull enough of copyright's harsh remedial levers to scotch the whole enterprise.

No, Yahoo! would need the same magic device of the class action that Google is now taking advantage of. Would the plaintiffs bother to organize themselves as a class for its benefit? There's no guarantee they would. Yahoo! would be in the especially tricky situation of filing a declaratory judgment action against a class of copyright owner defendants. It'd be hard even just to pick proper class representatives and appoint appropriate class counsel without some kind of collusion. And once there were class representatives, would they settle on comparable terms? There's no guarantee of it, especially given the guaranteed good deal that they're getting from Google.

I've also heard floated the idea that competitors are perfectly free to lobby Congress on orphan works legislation. So they are, but the argument that lobbying is an acceptable substitute for free competition in the book market as it currently exists is laughable. If I manufacture widgets and my competitor is monopolizing the widget market, it's no answer to my pleas to say that I can ask Congress for widget subsidies. Orphan works legislation, done right, would be a great thing. But no one should have to count on it happening as a condition of entry to a market that Google is already in.

Thus, Google's first-past-the-post status here could easily turn into a durable monopoly. That might be the inevitable result anyway; this is a market with substantial economies of scale and positive network effects. One may or may not think that a book-search and distribution monopoly built on such structural bases is legitimate; one may or may not favor government intervention if it just so

happens that Google is the only player in this game. I take no position on these questions. But the court reviewing this settlement should not set up its own power—in the form of its ability to bind absent class members—as a barrier to entry in the online-books and book-search markets. What, then, should the court do?

- **No most-favored-nation clause.** The most pressing problem is that the settlement explicitly guarantees Google a privileged position, via a most-favored-nation clause in Google's favor. For 10 years, the Registry can't give anyone else better "economic and other terms" than Google gets.²⁸ Notice, for example, that this term would preclude the Registry from offering a better revenue-sharing deal to Yahoo! even if the Registry thinks that this better deal is necessary to turn Yahoo! into a serious competitor to Google. This clause alone might be enough to deter any other serious entry. Google's concern about being undercut is real, but provided that the Registry itself is under proper antitrust scrutiny (see above), it has nothing legitimate to fear. The most-favored-nations clause should be struck.
- **Competitors offered the same deal as Google.** Any other entity willing to assume the same payment and security obligations that Google assumes in the settlement should be allowed to offer the same services that Google will, or any subset of them. This kind of competition is fair to authors and publishers because the various payment and security terms are already presumably acceptable to them. It's fair to Google, which gets the same deal it currently does. It's fair to competitors, who could enter on a level playing ground, without needing to roll the dice on invoking the legal system's power to intervene. And it saves the legal system the work of having to deal with the Microsoft Book Search lawsuit, the Yahoo! Book Search lawsuit, the Facebook Book Search lawsuit, and so on.
- **Registry authority to negotiate with Google competitors.** The previous rule suffices for the services the settlement describes, but there are also the new business models, which can't be specified in detail precisely because they don't exist yet. Under the proposed settlement, the Registry can give its blessing to Google on plenty of projects; it should be allowed to give the same blessing to anyone else. Crucially, that blessing would have the same effect of binding all authors and publishers to the deal it strikes. Again, authorizing the Registry to do such things is fair because copyright owners could still opt out of any new uses.²⁹ Once the Registry did bless a project—by

Google or by a competitor—the previous rule would kick in and require that the new project license be available to anyone on nondiscriminatory terms.

- **Beware a scanning monopoly.** If Google remains the dominant player in actually scanning books, we ought to be concerned about steps that it takes to preserve or extend that monopoly.³⁰ This concern is especially acute for public-domain books, which ought not to be under anyone's exclusive control. Google's Web search engine has the right model: It returns results from Google-run sites alongside results from sites run by others. Google's book search engine should be similarly ecumenical and treat book scans put online by others on an evenhanded basis. Google should also be prohibited from using its terms of service to put copyright-like terms-of-service restrictions on what others do with its public-domain scans.³¹ In the opposite direction, Google's collection of public-domain books should be open to other search engines.

PRINCIPLE 3: CONSUMERS NEED PROTECTION

Despite being the product of a lawsuit in which the reading public isn't directly represented, the proposed settlement often takes the public interest seriously. Many clauses in it aren't strictly necessary to resolve the dispute between the parties to the lawsuit but nonetheless go a good way toward making sure that the results will provide books to the public on fair terms:

- Google specifically promises that it won't use pop-up or pop-under ads, and the Registry is also authorized to take swift action to opt authors out of having their books shown with "animated, audio or video advertisements."³²
- In providing the Institutional Subscription, Google guarantees that its terms and conditions will "not prohibit any uses . . . that would otherwise be permitted under the Copyright Act."³³ This is significant; it's a commitment that the subscriptions won't require readers to surrender their fair use rights, for example.
- Similarly, the Institutional Subscription will never offer an "experience and rights" worse than those enjoyed by readers who purchase e-books through the program,³⁴ and copyright owners won't be allowed to sell out-of-print books while excluding them from the Institutional Subscription.³⁵
- Colleges and public libraries will receive free computer terminals with complete access to the Institutional Subscription.³⁶ The service is stingy by default: Four-year colleges will receive one terminal per 10,000

students, and public libraries one per building. Google isn't even obligated to provide the service at all. (There's authority for Google and the Registry to expand this program, but it's unclear that the Registry would ever approve more generous terms.)

- Various provisions of the Copyright Act operate in favor of libraries and other public-service entities, and the settlement makes sure that those promises remain intact. Libraries that allowed their books to be scanned get back digital copies, which they can use for accessibility purposes and to replace damaged or lost copies of physical books.³⁷

These provisions are all to the good. There are, however, other consumer-protection matters on which the settlement is silent or ambiguous. Google insists on its good intentions in many of these areas. ("Don't be evil," and all that.) I would prefer to see stronger protections than Google's continued promises of non-evilness. It's in the nature of such assurances that they're still offered long after they've ceased to be true. To the extent that Google really means to abide by them, it should have no objection to putting these terms explicitly in the settlement agreement or in an FTC consent decree.

- **No price discrimination.** The pricing structure for the Institutional Subscription will be largely fair. (Google will set pricing on an FTE basis with different pricing buckets for different categories of institutions, for example, higher education, corporate, government, etc.³⁸) For individual buyers, either booksellers can pick a price, or they can let Google's algorithms set a price for them based on buying patterns.³⁹ The settlement doesn't explicitly say that Google won't charge different readers different prices for the same book, something that its immense computational power and huge pricing corpus might make dangerously attractive. Google has no current plans to do so, and it's true that the market punished Amazon harshly when it tried the same stunt a few years back, but still. Better safe than sorry.
- **Reader privacy.** Your choice of reading matter ought to be highly private. There's a real concern that Google could identify and track readers, page by page, minute by minute. Indeed, the security standard that Google must comply with requires it to keep extensive logs of user activity.⁴⁰ Similarly, libraries that open their digital copies for scholarly and classroom uses must "keep track of and report[] all such uses of Books to the Registry."⁴¹ The only explicit privacy protections in the settlement, though, are about keeping private the information that the Registry has

about copyright owners.⁴² That's insufficient. The settlement should contain explicit privacy guarantees that user information and reading habits should be monitored only to the minimal extent necessary for billing, auditing, and security; that no such data be used for any other purpose, that all such data be promptly destroyed when no longer needed; that Google not reveal any information about any user or users' reading habits to any other entity, including the Registry; and that Google be legally responsible for any security breaches resulting in third-party access to reader information.

- **Reasonable terms and conditions.** The settlement should protect reader rights under the Copyright Act across the board. There's a good first cut at such language in the settlement already. Google promises reasonable terms and conditions for the Institutional Subscription, specifically that it will "not prohibit any uses . . . that would otherwise be permitted under the Copyright Act."⁴³ Similarly, the Library-Registry agreements should contain explicit statements that library terms of service will not require readers to give up any of their other rights under the Copyright Act. Any new business models should come with the same protections, as well.

PRINCIPLE 4: PUBLIC GOODS SHOULD BE WIDELY AVAILABLE

The Google Book Search project would be impossible without some crucial bibliographic databases. A pile of scans is useless unless it's linked to a database of publication metadata. Since only out-of-print books will be previewable by default, Google needs a database telling it which books are in and out of print.⁴⁴ To convey copyright-owner requests to Google and to convey payments from Google to them, the Registry will need a database of book copyright ownership. Some of these databases exist already; some will be built or supplemented as part of the project.

These databases are all public goods. They'll be useful to readers and researchers. They're also going to be immensely useful to players in the book business. The in-print database will help libraries understand their rights under copyright law; the rights-owner database will help publishers gather the rights that they need to publish new and exciting editions.

Moreover, these databases are byproducts of the Google Book Search project, not its goals. Google isn't compiling them because it can make money selling access to the databases. Instead, it's compiling them because it can't offer Book Search without them. Whether or not

Google (or the Registry) can monetize these databases directly won't substantially affect the incentive to compile them.

Taken together, these propositions imply that these databases should be opened to broad public access. That's exactly the policy that the settlement takes with the Books Database—Google's list of books it has or plans to scan—which is required to be online and searchable.⁴⁵ The same policy should be adopted wherever else possible.

That's not everywhere. I'm informed that Google has assembled its database of bibliographic metadata about publication largely by licensing it from other sources. Scholars may contest whether such databases should be capable of exclusive licensing, but even those who want to pick that fight shouldn't pick it here. Google didn't generate this data; it shouldn't be forced to reveal it.

The correct principle, instead, is that to the extent that Google and the Registry create new and useful metadata databases as part of the Book Search project, those databases should be offered to the public, gratis, and without legal or technical restrictions. The settlement agreement contemplates at least two such databases, both important (though the principle might also apply to others).

- **Publicly available in-print information.** Consider first the database of in-print information that Google needs to decide whether works are “commercially available” and thus restricted by default. Google currently synthesizes this information from a variety of sources (such as looking at used book sales online).⁴⁶ Google is required by the settlement to make this information available to the Registry on behalf of copyright owners;⁴⁷ it should be required to make the database public, as well. This isn't likely to be a practical problem, since Google will all but inevitably expose this information when it lets users either see preview pages or not. But still, Google shouldn't be given the option to restrict its availability; in case Google does wind up having competitors in this space, there will be an inevitable temptation to cut off access as a way of slowing down the other guy.
- **Publicly available copyright ownership information.** There's also the database of information about copyright claims that the Registry will need to use to distribute payments among copyright owners. The Registry is required to share much of this information with Google;⁴⁸ it should be required to share almost as much of the copyright-owner database with the public. There are privacy concerns here, since it will contain information about authors, but those concerns can be accommodated without much limiting the usefulness of the database in solving orphan works

problems. Pseudonyms and proxies are reasonable, provided that the database in general is made available so that others can use it as a point of contact in finding copyright owners or in verifying that no one knows who the owner is or where she can be found.

- **Best practices for open access.** In making these databases available and in providing some of the other core services (the exact set to be determined), Google should be required to use standard APIs and open data formats, as well as to allow programmatic access and bulk download where appropriate. Google currently does this as a matter of policy in many of its other lines of business, and it's already providing PDF downloads of public-domain books. These good policies should be enshrined as actual requirements. Among other things, they'll ensure that Google's competitors behave reasonably, too.

PRINCIPLE 5: ACCOUNTABILITY AND TRANSPARENCY MATTER

Google has been repeatedly criticized by scholars and activists upset at its lack of institutional transparency. It's also learned from that criticism. The settlement agreement contains some reassuring provisions to provide accountability. As good as they are, they should be supplemented with a few more.

Institutionally, Google and the Registry are given mutual rights to audit each others' relevant books.⁴⁹ While these audits are themselves confidential, the arrangement creates a healthy system of mutual accountability. Similarly, research users and libraries are subject to security audits under suitable procedures.⁵⁰ When there are disputes about public-domain or in-print status, they're subjected to a low-stakes initial process that lets the parties sort out the facts.⁵¹ Larger disputes go first through executive-level mediation and then arbitration, on reasonably balanced terms.⁵²

Google has also accepted a fairly stringent set of rules that prohibit it from altering the texts of the books that it scans. There are the usual, sensible carve outs: Google can hyperlink indices, it can link from books to sources they cite, it can highlight user search queries, and it can even add a limited social-networking annotation-sharing feature.⁵³ These are specific exceptions, however, from the general principle that it won't change one word of the author's writings without permission.⁵⁴ Good.

Google has even agreed to procedures that limit its editorial discretion to exclude books from being displayed. If Google removes a book for “editorial reasons,” it will tell the Registry about it and give the Registry a digital copy of the book.⁵⁵ The Registry may then go out and commission

a competitor to provide the display services that Google has refused to.⁵⁶ Google believes it has a First Amendment right not to be required to “speak” by passing along a book that it strongly objects to, and it’s chosen an honorable and speech-friendly way of exercising that right. Google’s waiver does not censor the book itself, which can still be made available through other means.

There are, however, some potential accountability holes in this system. One is that the Registry need not, or might not be able to, find a replacement for Google. I’m reluctant to intervene too strongly here, particularly when no other potential partner is willing to step forward. Others, bolder than I, might propose a positive duty on the Registry’s part, but I take no position on the issue, noting only that it raises difficult issues of free speech law and free speech policy. Fortunately, other potential holes are easier and less controversial to close.

- **No secret censorship.** If Google de-lists a book and the Registry doesn’t or can’t engage a replacement, the book will genuinely vanish from this new Library of Alexandria. Perhaps that should happen for some books, but decisions like that shouldn’t be made in secret. When Google chooses to exclude a book for editorial reasons, it should be required to inform the copyright owner and the general public, not just the Registry. This path leaves intact Google’s option to be silent, but requires that it be exercised with transparency. If and when Google chooses not to speak, it should own the ethical consequences, rather than being able to hide from its decision to hide a book.
- **Clear definition of non-editorial exclusion.** The settlement contains no clear distinction between “non-editorial” and “editorial” reasons for Google to exclude a book from being displayed. This ambiguity raises the possibility that Google might exclude a book for editorial reasons but tell no one, not even the Registry, about it, and thereby completely suppress the book. There’s a danger of line-crossing wherever a line is drawn, but assuming that Google will act in good faith, a sharper definition of “non-editorial reasons” should suffice. The current draft of the settlement says “quality, user experience, legal, or other non-editorial reasons,” an unclear and imprecise list that could easily be converted into a clear and precise one.
- **Accurate scanning.** Under the settlement, Google can’t “intentionally alter” the text of the books that it scans,⁵⁷ but it doesn’t promise anything about the quality of the digitization process itself. Unfortunately, scanning mistakes have left some of its digital books all but unreadable.⁵⁸ Google should be required to

institute a review program that would take reports of distorted or mutilated digital versions, respond with reasonable speed, and rescan the books if necessary (and allowed by the owner of the physical copy).

- **Other scanning institutions.** One last point of accountability concerns an issue raised by Jean-Noël Jeanneney: What books are scanned and in the collection at all?⁵⁹ Jeanneney’s specific concern—a lack of Francophone sources—has an easy and obvious response: The Bibliothèque nationale de France, of which he is the president, could join with Google to scan its collections. Indeed, Google has indicated its broad willingness to partner with libraries interested in scanning large corpuses of books to get them into the digital collection more quickly. Once again, Google’s sensible policy is one thing in the context of a private Google project and another in the context of a massive remaking of the US system of book copyrights that requires the blessing of a court of law. So long as Google is the only serious player in book scanning and search, any institution that wishes to provide books for scanning, or to perform scanning itself, should be allowed to take part in the scanning effort and ensure that particular works are digitized. There will need to be appropriate provisions about capacity, financing, quality control, and so on, but a well-drafted consent-not-to-be-unreasonably-withheld clause can take care of many of them. Once again, it’s worth emphasizing that this provision, like all of the others, would apply both to Google and to any of its competitors who come in under the modified settlement.

CONCLUSION

The starting point for my analysis has been that Google and the copyright owners are asking a federal court to put the US judicial power behind a document that they have presented to it. The court’s consent should not be given lightly; the settlement should be approved only when the court is satisfied that it really will serve the interests of all parties, including the public. I have tried to offer general principles to think through what the public interest requires, along with specific, realistic recommendations to implement those principles.

At the same time, this is not a sentencing hearing or a legislative chamber. The court is not in a position to rewire Google and the book industry to right all wrongs therein, nor should it try. Google’s other ventures are not on the table; nor are the many other problems bedeviling copyright law. My recommendations respond to the specific question that the court faces: Should it use its

power to bind absent class members and approve this settlement?

Thus, I hope that my recommendations all have two things in common. Each takes off from some issue specifically raised by the proposed settlement, some way in which approving the settlement could cause trouble down the line. Each then offers a change to head off that trouble, a change more or less narrowly tailored to the issue it confronts.

How do we get there from here? The parties can't settle the case without the court's approval, but the court can't just rewrite the settlement and impose it on them, either. The best alternative would be for Google, the authors, and the publishers to modify the proposed settlement along the lines that I've suggested and then bring the modified version to the court for approval. None of my recommendations touches the basic deal at the heart of the settlement or seeks to impose terms that the parties should find onerous.

We don't need to depend on the parties' good graces, however. Some of the concerns that I've discussed can easily be raised by members of the plaintiff class as part of an objection to the fairness of the settlement to them. Authors have a particular interest in not having their work unaccountably excluded from Google Book Search, for example. They should enter objections to the settlement unless it's modified. Amicus briefing on some of these issues could help the court put the settlement in context, as would intervention by some of the many interested non-parties.

Even if the court is unwilling to fit all of these issues into the copyright dispute before it, there are other legal avenues open. The DoJ could open an investigation into the antitrust issues and the Federal Trade Commission into the consumer-protection ones. It matters less that these recommendations be embodied in the settlement than that they be enshrined somewhere enforceable, and consent decrees are a perfectly reasonable alternative. Potential competitors in book-scanning can also raise many of these issues, for example, by way of a private antitrust suit; Microsoft, which previously had a book-scanning project, would be a natural plaintiff.

My goals here are pragmatic. I'm not proposing to take public control of the Book Search project. In comparison with the institutional reconfiguration of book copyright law that the settlement would enact, these tweaks are all quite minor. Nor am I proposing to leave Book Search entirely alone; the parties gave up on that possibility when they asked the court to approve this sweeping class-action settlement.

I hope that these recommendations will prove equally appealing to those who think that Google can do no evil

and those who think that it does only evil. Perhaps they'll prove equally frustrating. I offer them not as criticisms of the settlement, but refinements of it. As the chess adage goes, "When you see a good move—wait—look for a better one!"

NOTES

1. See Proposed Settlement, *Authors Guild v. Google Inc.*, No. 05 CV 8136 (S.D.N.Y. filed Oct. 28, 2008) [hereinafter Settlement]. The case has been coordinated with and would be settled along with the related publisher lawsuit, *McGraw-Hill Cos. v. Google, Inc.*, No. 05 CV 8881 (S.D.N.Y. complaint filed Oct. 19, 2005).
2. Settlement art. X. I
3. Settlement art. V.
4. Settlement §§ 3.1(a), 3.3(a)–(c), 3.4.
5. Settlement §§ 1.86, 4.5(a)(ii).
6. Settlement art. VI.
7. Settlement § 4.1.
8. Settlement § 4.8.
9. Settlement § 4.2.
10. Settlement § 4.2(b)–(c).
11. Settlement § 4.5(a)(i).
12. Settlement § 7.2(b)(vi), (d).
13. Settlement § 7.2(d)(vi)(1).
14. Settlement § 7.1.
15. Settlement § 7.2, attachment B1.
16. Settlement attachments B2, B3.
17. Settlement § 3.14.
18. Settlement art. VI.
19. Settlement § 8.2(a)(iv).
20. Settlement attachment C § 1.1(a).
21. Settlement § 4.7.
22. Settlement § 6.2(b).
23. Settlement § 6.2(b).
24. Settlement § 2.4.
25. Settlement § 3.5(a)(iii).
26. Settlement § 3.5(b)(i).
27. Settlement §§ 1.16, 1.38, 1.142. This statement leaves out copyright interests in foreign works; under US law, one need not register such a work before suit. See 17 U.S.C. § 411(a). Owners of unregistered copyrights in foreign works are part of the plaintiff class and thus part of the settlement.
28. Settlement § 3.8(a).
29. Settlement § 4.7.
30. See Tim O'Reilly, "Book Search Should Work Like Web Search," O'Reilly Radar (Dec. 11, 2006), <http://radar.oreilly.com/archives/2006/12/book-search-sho.html>.
31. See Yakov Shafranovich, "Change in Google Book Search Guidelines for Public Domain Books" (Jan. 30, 2009), <http://www.shaftek.org/blog/2009/01/30/change-in-google-book-search-guidelines-for-public-domain-books/>.
32. Settlement § 3.10(c)(iii).
33. Settlement § 4.1(e).
34. Settlement § 4.1(f).
35. Settlement § 3.5(b)(iii).
36. Settlement § 4.8.
37. Settlement § 7.2(b).
38. Settlement § 4.1(a).
39. Settlement § 4.2(b)–(c).

-
40. Settlement attachment D § 3.
 41. Settlement § 7.2 (b)(vii)
 42. Settlement § 6.6(a)(vi), (d).
 43. Settlement § 4.1(e).
 44. Settlement § 3.2(b).
 45. Settlement § 3.1(b)(ii).
 46. Settlement § 3.2(d).
 47. Settlement § 3.2(d)(ii).
 48. Settlement § 6.6(c).
 49. Settlement § 4.6(e), 6.3(d).
 50. Settlement § 8.2(c).
 51. Settlement § 3.2(d)(iv)–(v).
 52. Settlement art. IX.
 53. Settlement § 3.10(c)(ii).
 54. Settlement § 3.10(c)(i).
 55. Settlement § 3.7(e).
 56. Settlement § 3.7(e)(i).
 57. Settlement § 3.10(c)(i).
 58. See Paul Duguid, “Inheritance and Loss? A Brief Survey of Google Books,” *First Monday* (Aug. 2007), <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1972/1847>.
 59. See Jean-Noël Jeanneney, *Google and the Myth of Universal Knowledge* (2007).