Legally Speaking
The Dead Souls of the Google Book Search Settlement

Why the Google Book Search settlement agreement under consideration could result in an extensive restructuring of the book industry.

Google has scanned the texts of more than seven million books from major university research libraries for its Book Search initiative and processed the digitized copies to index their contents. Google allows users to download the entirety of these books if they are in the public domain (about one million of them are), but at this point makes available only "snippets" of relevant text when the books are still in copyright (unless the copyright owner has agreed to allow more).

In the fall of 2005, the Authors Guild, which then had about 8,000 members, and five publishers sued Google for copyright infringement. Google argued that its scanning, indexing, and snippet-providing was a fair and non-infringing use because it promoted wider public access to books and because Google would remove from its Book Search repository any digitized books whose rights holders objected to their inclusion. Many copyright professionals expected the Authors Guild v. Google case to be the most important fair use case of the 21st century.

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 column title refers) would agree that the settling authors and publishers will truly represent their interests when setting terms for access to Book Search.

Orphan Works
An estimated 70% of the books in the Book Search repository are in-copyright, but out of print. Most of them are, for all practical purposes, "orphan works," that is, works technically still in copyright, but for which it is virtually impossible to locate the appropriate rights holders to ask for permission to digitize them.

A broad consensus exists about the desirability of making orphan works more widely available. Yet, without a safe harbor against possible infringement lawsuits, digitization projects pose significant copyright risks. Congress is considering legislation to lessen the risk of using orphan works, but it has yet to pass.

The proposed Book Search settlement agreement solves the orphan works problem for books—at least for Google. Under this agreement, which must be approved by a federal court judge to become final, Google would get, among other things, a license to display up to 20% of the contents of in-copyright out-of-print books, to run ads alongside these displays, and to sell access to the full text of these books to institutional subscribers and individual purchasers.

The Book Rights Registry
Approval of this settlement would establish a new collecting society, the Book Rights Registry (BRR), initially funded by Google with $34.5 million. The BRR will be responsible for allocating $45 million in settlement funds that Google is providing to compensate copyright owners for past uses of their books.

More important is Google's commitment to pay the BRR 63% of the revenues it makes from Book Search that are sub-
How can Google be getting a license to make millions of in-copyright books available through Book Search just by settling a lawsuit?

U.S. law allows the filing of “class action” lawsuits whose lead plaintiffs claim they represent a class of persons who have suffered the same kind of harm as a result of the defendant’s wrongful conduct as long as there are common issues of fact and law that make it desirable to adjudicate the claims in one lawsuit instead of many.

The Authors Guild and three of its members sued Google, claiming to represent a class of similarly situated authors whose books Google was scanning and whose copyrights Google was violating. By bringing a class action lawsuit, the Authors Guild put considerable financial pressure on Google because the winner of a class action lawsuit is entitled to an award that equals all of the monies owed to the class, which may be exponentially higher than awards to individual plaintiffs.

In the absence of a settlement agreement, Google would almost certainly have vigorously fought against certification of the class in the Authors Guild case. After all, the guild has only a few thousand members and most of them do not write the kinds of scholarly works that are typically found in major university research libraries. Many scholarly book authors might want their books to be scanned by the Book Search project so they will be more accessible to potential readers.

The publisher lawsuit did not start out as a class action lawsuit, perhaps in part because McGraw-Hill et al. recognized how difficult it would be for them to prove they adequately represented a class of all book publishers whose books Google had scanned.

However, the agreement that Google

Class Action Settlement

By now, Communications readers may be a bit puzzled. How can Google be getting a license to make millions of in-copyright books available through Book Search just by settling a lawsuit brought by a small fraction of authors and publishers?

subject to sharing provisions. The revenue streams will come from ads appearing next to displays of in-copyright books in response to user queries and from individual and institutional subscriptions to some or all of the books in the corpus. Google and the BRR may also develop new business models over time that will be subject to similar sharing.

One of the main jobs of the BRR will be to distribute these revenues. The money will go, less BRR’s costs, to authors and publishers who have registered their copyright claims with the BRR. Although the settlement agreement extends only to books published prior to January 5, 2009, the BRR is expected to attract authors and publishers of later-published books to participate in the revenue-sharing arrangement that Google has negotiated with the BRR.
has negotiated with the Authors Guild and the Association of American Publishers (AAP) would, if approved, be settled as a class action on behalf of all book authors and publishers, with the Guild and AAP claiming to represent their entire respective classes. By according to the certification of these classes through the settlement, Google will get a license from all authors and publishers of books covered by the agreement (which is to say nearly every in-copyright book ever published in the U.S.) so that it can commercialize them through the Book Search.

Google's New Monopoly
The proposed settlement agreement would give Google a monopoly on the largest digital library of books in the world. It and the BRR, which will also be a monopoly, will have considerable freedom to set prices and terms and conditions for Book Search's commercial services. The BRR is unlikely to complain that the price is too high, the digital rights management technology is too restrictive, or the terms are too onerous. Google will also be the only service lawfully able to sell orphan books and monetize them through subscriptions. The BRR will get 63% of these revenues that it will pay out to registered authors and publishers, even as to books in which they hold no rights. (Some unclaimed orphan work funds may go to charities that promote literacy.) No author whose books are in the corpus can get paid by the BRR unless he or she has registered it.

Virtually the only way that Amazon.com, Microsoft, Yahoo!, or the Open Content Alliance could get a comparably broad license as the settlement would give Google would be by starting its own project to scan books. The scanner might then be sued for copyright infringement, as Google was. It would be very costly and risky to litigate a fair use claim to final judgment given how high copyright damages may be (up to $150,000 per infringed work). Chances are also slim that the plaintiffs in such a lawsuit would be willing or able to settle on equivalent or even similar terms.

Dead Souls
The Book Search settlement brings to mind Nikolai Gogol's story, Dead Souls.

The Book Search agreement under consideration is not really a settlement of a dispute over whether scanning books to index them is fair use.

Chichikov, its main character, travels around the Russian countryside to buy "dead souls" in an attempt to become a wealthy and influential man. In the early 19th century, Russian landowners had to pay annual taxes on the number of serfs—counted as "souls"—they owned as of the last census.

Chichikov offered to buy "dead souls" (peasants who had died since the last census) from the landowners. His plan was to acquire enough of these souls so that he could take out a large loan secured by his portfolio, and thereby become a wealthy man.

In Gogol's story, Chichikov's scheme falls apart. Rumors fly that the souls he owns are all dead and he flees the town in disgrace. In Google's story, however, the dead soul scheme seems likely to pay off handsomely, as Google will have the exclusive right to commercially exploit millions of orphan books.

Representativeness?
As galling as it is to realize that the BRR and its registered authors and publishers will derive income from millions of books they didn't write or publish, it is even more galling that copyright maximalists will almost certainly dominate the BRR governing board.

(The Authors Guild president, for example, complained about the "read aloud" feature of Kindle, denoting it a "swindle," and a copyright infringement. The AAP is supporting legislation to forbid the U.S. National Institutes of Health from promoting "open access" policies for articles written under NIH grants. And of course, the Authors Guild and AAP characterized Google as a thief for scanning books from research libraries.)

If asked, authors of orphan books in major research libraries might want their books to be available under Creative Commons licenses or even be put into 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 term of access most authors of books in the corpus would want.

In reviewing the settlement, the judge is supposed to consider whether the settlement is "fair" to the classes on whose behalf the lawsuits were brought. He may assume the settlement is fair because money will flow to authors and publishers. But importantly absent from the courtroom will be the orphan book authors who might have qualms about the Authors Guild and AAP as their representatives.

Conclusion
In the short run, the Google Book Search settlement will unquestionably bring about greater access to books that major research libraries collected over the years. But it is very worrisome that this agreement, which was negotiated in secret by Google and a few lawyers working for the Authors Guild and AAP (who will, incidentally, receive up to $45.5 million in fees for their work on the settlement—more than all of the authors combined!), will create two complementary monopolies with exclusive rights over a research corpus of this magnitude. Monopolies are prone to engage in many abuses.

The Book Search agreement under consideration is not really a settlement of a dispute over whether scanning books to index them is fair use. It is a massive 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 in its current form without modification.

Pamela Samuelson (pam@law.berkeley.edu) is the Richard M. Sherman Distinguished Professor of Law and Information at the University of California, Berkeley.

Copyright held by author.