



# What's at Stake in the Google Book Search Settlement?

PAUL N. COURANT

In 2004, Google embarked on a project of historic scope. Its aim was to scan and index the contents of the world's great research libraries. One small problem: authors and publishers of works in copyright sued, arguing that Google violated copyright law by scanning works without the explicit permission of rightsholders. In the way of most commercial lawsuits, the parties settled, but unlike most commercial lawsuits, the settlement greatly expanded the stakes, creating a great electronic bookstore where Google would sell (with most of the revenue going

---

*Paul N. Courant is the Harold T. Shapiro Collegiate Professor of Public Policy and the Arthur F. Thurnau Professor of Economics and of Information at the University of Michigan. He is also the University Librarian and Dean of Libraries, and was Provost at Michigan when Google negotiated its original arrangement to digitize the University's collections.*

to rightsholders) access to millions of copyrighted works, something not contemplated in Google's original scanning project. The settlement engendered many howls of protest and some expressions of support. Pamela Samuelson, among the most eloquent of the objectors, called the proposed settlement "audacious" and argued that it is "designed to give [Google] a compulsory license to all books in copyright throughout the world forever."

In September, weeks before a scheduled fairness hearing that could have led to approval of the settlement by the court, the court asked the parties for an opportunity to revise the settlement agreement in light of concerns raised by the U.S. Department of Justice and hundreds of other objectors. The court has given the parties a November deadline for a revised agreement.

In my opinion, approval of the original settlement would have been vastly preferable to its rejection, because it provided extraordinary and valuable benefits to readers and scholars. But, in at least one important feature, I believe that the Department of Justice may be pointing the way to a settlement that would be better than the original. Before I come to these conclusions, some background is required.

## COPYRIGHT AND THE GOOGLE SCANNING PROJECT

Article I, Section 8 of the United States Constitution gives Congress the power to "promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries." The framers struck a classic balance in public

economics. They recognized simultaneously that the incentive to develop new knowledge would be enhanced if writers and inventors could capture some of the return on their work (hence the “exclusive” rights, which are not completely exclusive, with good reason). At the same time, they recognized that knowledge and new discoveries are nonrival, meaning that added users do not diminish the uses of existing users (hence the free use after “limited times”).

Congress has repeatedly extended the “limited time” that is covered by copyright, with the result that most works written in the twentieth century are protected by copyright and unavailable to be read on-line, even when a perfectly readable scanned copy exists. In brief, what’s at stake in the Google Book settlement is our ability to find and read the literature of (most of) the twentieth century on-line.

Of course, the great majority of works written in the twentieth century are out of print, and most have been out of print since shortly after their publication.

Until Google began its mass digitization of research library collections, access to

out-of-print works generally required visiting a library that had a copy. As the digitization project has proceeded (well over twelve million books have been digitized and indexed by Google, with tens of thousands being added every week) an interesting and instructive bifurcation has occurred.

Digitized works known to be in the public domain—including essentially everything published in the United States before 1923 and many works published between 1923 and 1963—are easily available. They can be found through Google Book Search ([books.google.com](http://books.google.com)) and on the websites of many participating libraries. They can be read on-line or downloaded, and in the past several months hundreds of thousands have been made commercially available through a variety of print-on-demand channels, as well as on e-book readers. However, works that may still be in copyright are not so easy to find and to use. Google shows only a brief snippet in response to an on-line search, along with information about where a physical copy of the book might be found. The reason for the restricted usability of these works is simple: the penalties for displaying significant portions

of copyrighted material can be severe—from \$750 to \$150,000 (no, that’s not a typo; there are four zeros after the fifteen) per violation, and each internet hit would constitute a violation. This is a risk that neither Google nor any library is willing to take.

As result, most of the scholarly and cultural record of the twentieth century cannot be accessed by the methods preferred by pretty much everyone who reads in the twenty-first century. A professor wanting to assign a substantial portion of a book written in, say, 1961 (and out of print for 47 years), is generally still required to put the book on reserve the old-fashioned way, even if the university library has a digitized copy that could be put on a server and delivered to students at zero marginal cost.

I have referred to books published after 1922 as books that ‘may be’ in copyright. The copyright law is nothing if not complicated, and among its interesting features is that the current copyright status of books depends on dates of authorship and in many cases on the details of registration and renewal of copyright. To complicate matters further, for many books (the number is surely over half a

million in the U.S., in the millions worldwide) the holder of copyright is not known and may not exist, and often neither a living author nor a still-extant publisher knows who holds the rights to a given work. Such works are generally referred to as ‘orphan works’ in the world of libraries and copyright lawyers, but they are really a particular kind of orphan—foundlings. Their parents (holders of copyrights in them) are unknown. The existence of orphan works of this kind greatly complicates the already complicated and often expensive effort involved in finding rightsholders and making arrangements with them for lawful use of copyrighted works.

#### THE LAWSUIT AND PROPOSED SETTLEMENT

Google announced its mass digitization project with libraries in 2004. In 2005, several authors and publishers of copyrighted works, respectively working with the Authors Guild and the Association of American Publishers, sued Google for unauthorized copying of copyrighted works. Google’s defense was to argue that its indexing of the works and displaying of snippets constituted a fair use under the copyright law. Some three years later,

the parties to the suit agreed to a settlement, with the plaintiffs claiming to represent a class that includes essentially all those who hold copyright to works in libraries in the United States. The settlement agreement cannot go into effect without Federal District Court approval, and is currently under revision as a result of concerns that have been vigorously expressed by many parties, including the Department of Justice.

Basically, the proposed settlement would create a market in electronically-available copies of out-of-print works that are plausibly in copyright. Google would provide free browsing access (usually 20 percent of a book would be viewable in a given search) and would sell permanent on-line access to complete versions on behalf of itself and a newly-created Book Rights Registry (BRR) that would represent the interests of rightsholders of works subject to the settlement. Google would also sell site licenses to institutions such as colleges and universities, enabling students and employees of those institutions to have access to the collections in much the same way that they now have access to electronic journals and databases purchased by their libraries. Google would

obtain 37 percent of the revenue from both the retail product and the site license, with the rest to be distributed to rightsholders.

The obvious benefit of the settlement is that it provides electronic access to many millions of works at one fell swoop, saving the transactions costs that would be involved if Google, libraries, and others seeking to provide access to the scanned works had to negotiate work by work and rightsholder by rightsholder, assuming that rightsholders could be found. The ability to search simultaneously the collections of the world’s great research libraries, to browse those collections and to be able to purchase immediate electronic access provides uncalculated, but almost certainly large, consumer surplus.

The settlement would also permit academic libraries and their universities, at least after a time, to save a great deal of money and space, as the necessity of holding extensively duplicated print collections would be eliminated. Moreover, the settlement includes the orphan (foundling) works, removing the risks that would otherwise attend to displaying works where rights are unknown, and adding to the value that would be available

to students, professors, and other users of Google's newly-created giant electronic bookstore. Inclusion of the orphan works is essential to creating an effective product for the academic market because without them neither Google nor anyone else could risk putting collections on-line without costly establishment of rights (or the lack thereof) book by book, and the resulting collection of out-of-print works would be seriously incomplete. One can never conclusively prove a book to be orphaned. There is always the possibility that a rightsholder will materialize, with a lawyer not far behind.

#### OBJECTIONS

The objections to the original settlement agreement are many and varied. One set involves privacy, which I do not consider here. Others involve a cluster of legal questions, notably involving technical issues of civil procedure surrounding class action lawsuits, of which I have little interest and less understanding. Still others have a good deal of economic content.

The settlement would plausibly bestow considerable market power on the BRR and Google. Although other parties could choose

to scan libraries' works and provide market access by contracting with the BRR, only Google would be authorized to use the Google scans in the marketplace, and for at least the foreseeable future, the Google scans constitute the overwhelming majority of scans, because no one else has been willing to commit the hundreds of millions of dollars necessary to scan the tens of millions of works that reside in the world's research libraries.

Many objections to the Google settlement are based on a concern that Google and the BRR will have monopoly power in the newly-developed market for electronic access to out-of-print copyrighted works. Some assertions in this regard are simply silly: Google will not have a monopoly on access to the works, which are still available in the source libraries. Nor will it have been granted a monopoly on electronic access to works with known rightsholders; such works may be rescanned and sold by other providers, including rightsholders themselves, should it prove to be the case that the market in these works is sufficiently profitable to attract entry.

The orphan works are another matter. The settlement gives Google rights to distribute

orphan works and provides for no mechanism that could practicably convey such rights to other entities. Moreover, unlike the situation when rightsholders are known, there is no way for copyright holders of orphaned works to contract with other parties for distribution of those works, because it is in the nature of orphan works that the holders of the rights are unknown. This leads the Department of Justice to be concerned that the settlement, as drafted, "appears to create a dangerous probability that only Google would have the ability to market to libraries and other institutions a comprehensive digital-book subscription" (U.S. Department of Justice, p. 24).

#### EXPLOITING THE ORPHANS

The disposition of both the orphan works and revenues attributed to them are among the most controversial features of the settlement. One important economic objection is distributional. There is clear public benefit to providing access to orphan works. But under the settlement agreement, if the rightsholders to orphan works are not found after five years, the principal beneficiaries of the revenues attributable to orphan works would be

the authors and publishers who hold the rights to other works. Why them? Well to be honest, one can't help thinking that it is because these rightsholders were active parties in the lawsuits whereas orphans, parents, consumers and others were not. Other obvious possibilities would be to return these funds to consumers as rebates, to give them to the libraries that bought and preserved the works in the first place (would I have thought of this were I not Michigan's librarian?), or to reduce the federal deficit.

That Google has an advantageous position with regard to the orphan works is clear. Whether that position is worth any money to speak of is an open question. Books that are in copyright and out of print have already proved themselves to be something less than star performers in the marketplace. To be sure, libraries will find it valuable to have access to as complete a collection as possible. At the same time, the settlement adduces a principle of 'broad access,' in determining prices for institutional licenses. Moreover, the orphan works will continue to be available in print form and as part of Google's retail product, and 20 percent of each book will be available for free preview.

Additionally, it seems likely that Google is more interested in attracting people to its site than it is in profiting directly from sales of books, and hence would prefer prices to be low.

Sadly, I don't have an estimate of the price elasticity of demand for out-of-print works with unknown owners, but I am confident that Google's ability to exploit its monopoly position in orphan works is far weaker than the ability of commercial journal publishers to exploit their position in fields where getting the next essential grant requires immediate access to the most current published papers. The fear among some librarians, including Robert Darnton of Harvard, that Google will be able to set extortionate prices in the manner of many commercial journal publishers (not including bepress), seems to me to misapprehend important differences between old books and current scientific journal articles. Professor Darnton and I have debated the question at some length in a recent exchange in the *New York Review of Books*.

Fortunately, concerns about legal barriers to entry can be laid to rest by changing the treatment of the orphan works while preserving the benefits of the settlement to the public and to

the scholarly community. All of the efficiency gains from the settlement's treatment of orphan works are due to the settlement's waiver of legal liability from their sale and display. That waiver makes the books available and generates value for those who would search and read them. The distribution of revenues to rightsholders of other books serves no compelling distributional interest.

Thus, any mechanisms that allowed other entities (e.g. libraries, publishers, the Internet Archive, Microsoft) to display and facilitate search of orphan works on the same or similar terms that Google has under the draft settlement would realize the efficiency gains without distributional harm. And by allowing entry, any such mechanism would eliminate the most troubling element of potential monopoly that could arise from the settlement.

Thus a revised settlement (as suggested by the U.S. Department of Justice, p. 25) that provided competitors with the ability to use the orphan works on the same terms as Google, or legislation with similar consequence, would be an unambiguous improvement over the original settlement. The great benefit of a new market in electronic versions of the literature of

most of the twentieth century could be realized with reduced risk of monopoly and without enriching unrelated parties with the fruits of orphans' labor. In contrast, scuttling of the settlement or greatly limiting the volume of works to be covered would put us back to where we started—the only people with good access to the scholarly and cultural record of the twentieth century would be people with physical access to research libraries, and even for them, that literature would be more difficult to access than most works published before or since.

---

Letters commenting on this piece or others may be submitted at <http://www.bepress.com/cgi/submit.cgi?context=ev>.

#### REFERENCES AND FURTHER READING

**Band**, Jonathan (2008) "A Guide for the Perplexed: Libraries and the Google Library Project Settlement," Association of Research Libraries. Available at: <http://wo.ala.org/gbs/alaarl-summary-document>. (A link to the settlement itself can be found at the same URL.)

**Darnton**, Robert (2009) "Google and the Future of Books," New York Review of Books:

56(2): February 12. Available at: <http://www.nybooks.com/articles/22281>. (See also Paul Courant's response and Darnton's rejoinder. Available at: <http://www.nybooks.com/articles/22496>.)

**Samuelson**, Pamela (2009) "The Audacity of the Google Book Settlement," The Huffington Post, August 10. Available at: <http://www.huffingtonpost.com/pamela-samuelson/the-audacity-of-the-google-book-settlement>. (See also links to other commentary by Professor Samuelson.)

**U.S. Department of Justice** (2009) "Justice Department Submits Views on Proposed Google Book Search Settlement." Available at: [www.usdoj.gov/opa/pr/2009/September/09-opa-1001.html](http://www.usdoj.gov/opa/pr/2009/September/09-opa-1001.html). (Case 1:05-cv-08136-DC Document 720. Filed 09/18/2009.)

**American Library Association**, Available at: <http://wo.ala.org/gbs>. (Google Book Search Settlement website has many of the basic documents and a good deal of commentary, including a list of parties who filed comments with the court.)

