Should private companies like Google be able to build a digital library by scanning millions of copyrighted books?

Issue Date: August 8, 2011

SUPPORTERS ARGUE

If the goals set for Google's digital library are reached, it will offer scholars, students and average readers unprecedented access to millions of obscure, out-of-print, and forgotten texts. Furthermore, since anyone with access to the Internet will be able to find, buy, or borrow scanned books, the library will serve as a great equalizer, allowing all those who do not have physical access to great university and public research libraries to use those libraries' resources electronically. By bringing long-forgotten and out-of-print books back to life through digitization, the Google project will help the publishing industry by creating a market for previously low-value books.

OPPONENTS ARGUE

Google's book search project violates copyright law by scanning and indexing copyrighted work without the explicit permission of the authors or publishers who hold those rights. Putting such a vast digital library in the hands of a private company threatens to privatize the library system and leaves information vulnerable to censorship. A large-scale digital library should be a public venture, with the participation and oversight of the federal government and national public libraries.

In 2004, Internet giant Google announced plans to create the world's largest digital library by scanning millions of books and indexing them online. To launch the project, Google formed partnerships with five of the world's most prominent research libraries—the New York Public Library in New York City, Harvard University Library in Cambridge, Massachusetts, Stanford University Library in California, University of Michigan Library in Ann Arbor

An Internet archivist scans a book to be included in a digital library.
AP Photo/Ben Margot
and the University of Oxford Library in Great Britain—to scan their collections of books, many of which are obscure and out of print. Google hopes that soon Internet users can use Google Book Search to find, preview and buy or borrow any book in the world.

Currently, users can search the full texts of copyrighted books that have already been scanned but can view only small portions of the text, called "snippets." Google Book Search users can also view the full text of "public domain" works—books published before 1923 and not subject to copyright laws, which may be freely read and copied by anyone. In addition to its library project, Google also scans books through the "Partner Program," whereby authors and publishers can use Google Book Search to promote their works online by giving the company permission to scan and index their books. If rights holders grant permission through the Partner Program, Google can show longer "previews" of those books, including entire pages or chapters. Next to book previews, Google also provides links to online vendors, such as Amazon, where readers can purchase a copy of the book, or they can purchase Google's electronic version.

Google's ambition is to digitize all books ever published and make them available to scholars and average readers to buy or read online. In the company's words, the book project intends "to work with publishers and libraries to create a comprehensive, searchable, virtual card catalog of all books in all languages that helps users discover new books and publishers discover new readers."

While many observers have rejoiced at the prospect of an international, digital library at the world's fingertips, others have urged caution, fearing that Google will trample on authors' copyrights while essentially privatizing the public library system. Furthermore, in an uncertain publishing market, where sales of print books have been declining for years, publishers worry about the effect of large-scale digitization on the already shaken industry.

In 2005, a coalition of authors and publishers filed a class action lawsuit against Google, saying the company had committed copyright infringement by scanning millions of books from libraries without rights holders' expressed permission. (In a class action lawsuit, a small group of people or groups file a lawsuit on behalf of a much larger group of people who would also have an interest in the lawsuit and who usually share in any settlement.) Over the next few years, Google negotiated a compromise with litigants, finally agreeing in 2008 on a $125 million settlement. In March 2011, however, New York Second Circuit Court Judge Denny Chin rejected the terms of the settlement, saying it would violate antitrust laws by granting Google monopolistic advantages over the online book market. Chin also reprimanded the company for its "wholesale, blatant copying" of copyrighted works.

Critics of Google's ambition to digitize every book ever published rejoiced at the decision, while supporters lamented the loss, which overturned years of painstaking negotiations. Google, however, will likely continue to pursue the project by working out another settlement that can gain the court's approval. Does the Google Book Search project violate U.S. copyright law? Would it benefit readers and authors by expanding the market and increasing access to obscure works, or would it impose a monopoly, violating antitrust laws?

Critics say that Google's book project violates both copyright and antitrust laws. They also contend that it would put power over vast amounts of information into the hands of one private company. It would be safer, they say, for public institutions, such as national libraries, to undertake a national digital library project themselves. The U.S. government, not Google and a handful of litigants, should determine the terms of such a large-scale digitization project, critics say. Furthermore, opponents contend that Google's digital library would stunt competition in the digital book marketplace, handicapping an already struggling publishing industry.

Supporters of the Google book project, on the other hand, argue that by making millions of obscure and out-of-print works available to Internet users, Google would be providing an invaluable service to students, readers, researchers and scholars who may not live near one of the nation's major public research libraries. Google has the unique ambition and resources to embark on such a large-scale digitization project and should not be penalized for its initiative, supporters say. Furthermore, proponents note, digitization would preserve out-of-print obscure works, such as privately published memoirs, that could be invaluable to scholars but now remain susceptible to physical damage. Lastly, supporters argue that the Google book project would help the publishing industry by making rare
works more available to readers and scholars, thus creating a market for books that have long languished in obscurity.

**Google Embarks on Book Digitization Project**

Book digitization projects are not unique to Google. Since 1971, a volunteer digitization effort called Project Gutenberg has scanned thousands of public domain books and texts, which it then offers for free on its website for readers either to view online or download as electronic books, known as e-books. Most major U.S. research libraries and some university libraries have digitized at least some elements of their collections, as the Library of Congress did in the 1990s with a digital collection called "American Memory." Furthermore, advocacy groups consisting of authors, librarians and readers, such as the Cambridge, Massachusetts–based Open Knowledge Commons, have taken steps toward planning a national digital public library.

Digital library initiatives in other countries, however, have been far more ambitious than any public undertakings in the U.S., with many institutions making digital collections available for anyone in the public to view online for free. According to *New York Times* journalist Natasha Singer, "The U.S. finds itself trailing Europe and Japan in creating...a national digital library that would serve as an electronic repository for the nation's cultural heritage."

For example, in 2008, the European Commission—the executive body of the European Union (EU)—launched Europeana, a digital repository of 15 million artworks, books, music and videos from museums and other cultural institutions in EU member nations. In 2009, the United Nations Scientific, Educational and Cultural Organization (UNESCO) launched the World Digital Library, which seeks to digitize portions of collections from libraries in all U.N. member nations. Furthermore, many individual countries have already taken steps toward digitizing much of their national libraries' collections. The National Library of Norway, for example, announced in 2005 that it would digitize its entire collection, as did the National Library of the Netherlands in 2010. Proponents of a public digital library in the U.S. have pointed to the European models as alternatives to the privately run Google project.

Google, however, plans to surpass all previous book digitization projects. In 2002, the company's co-founder Larry Page devised a makeshift scanner at Google's headquarters to see how quickly it could scan books. Through partnerships with libraries and the development of rapid book scanning technology, the company digitized an estimated 15 million books by October 2010. From earlier, less ambitious plans, the company eventually announced its goal of creating a universally accessible digital library that would double as a bookstore. The project has reportedly become a favorite among employees of the company, which has a corporate mission to "organize the world's information and make it universally accessible and useful."

Under Google's plan, consumers could purchase online digital copies of millions of books. Furthermore, universities and libraries could buy subscriptions to Google's digital collections, so their students and researchers could utilize digital versions of texts from the nation's top public and university libraries.

Over the last decade, Google, whose informal corporate motto is "Don't be evil," has become an increasingly popular and omnipresent company in the digital world through its free search engine and e-mail services. Its book project has, however, sparked a severe backlash. Siva Vaidhyanathan, an associate professor of media studies at the University of Virginia in Charlottesville, told the *New York Times*, "This was the first issue through which Google's power became clearly articulated to the public. All sorts of people—writers, researchers, librarians, academics and readers—really feel they have a stake in the world of books."

In 2005, a group of authors and publishers, including the Authors Guild, a New York City–based authors' advocate group, filed a class action lawsuit against Google. The lawsuit contended that Google's project infringed on authors' and publishers' copyrights.

In its defense, Google said that previous court cases had set precedents establishing that the indexing of copyrighted material for online purposes was "fair use"—a legal doctrine that allows certain types of copyrighted material to be
reproduced, depending on the circumstances, in full or in part, without permission. According to the U.S. Copyright Office, fair use material includes copyrighted matter that may be fairly reproduced, "such as criticism, comment, news reporting, teaching, scholarship, and research." In order to differentiate fair use from infringement, according to copyright law, courts must analyze "the purpose and character of the use," such as whether it is intended for commercial or educational purposes, the nature of the copyrighted work, the "amount and substantiality" of the work used and "the effect of the use upon the potential market for, or value of, the copyrighted work." [See Copyright and the Internet]

Much of Google's "fair use" argument revolved around the contention that its digital library would be primarily an educational tool used to advance research and scholarly work in all fields. Part of its project, for example, was the Google Book Ngram Viewer. Unveiled in December 2010, the online tool allows users to track the frequency with which words or phrases are used in published books over time. [See Google Unveils 'Ngram Viewer' as Part of Controversial Book Project (sidebar)]

Google also argued that its "fair use" of book snippets would encourage readers to buy or borrow books that previously did not sell well. That contention raised a relatively new question in the arena of copyright law. According to the Congressional Research Service, much of the lawsuit addresses the question "Can digitization itself claim to be fair use on the grounds that" digitized works "facilitate legitimate uses"? In other words, can Google—and could other companies—be excused for scanning works without permission because the end result might lead to greater exposure and sales for those copyrighted works?

In addition to fears of copyright infringement, much opposition to Google's Book Search project concerns its "opt-out" model. Under that framework, Google would include material from authors and publishers in its digital collection unless those parties actively refused to participate. Many critics of the Google project have urged the company to adopt instead an opt-in model, so only copyright holders who actively wanted their works to be part of the project would be included.
Judge Rejects Google's Lawsuit Settlement

Complicating Google's initiative is the presence in its digital collection of so-called "orphan works"—books whose author or copyright holder cannot be located. For some of those books, the author has died or disappeared, the publishing company disbanded or was unresponsive, or the original publishing contract has been lost. Many orphan works housed in public libraries, such as privately published memoirs, are invaluable to historians and researchers. According to the Washington Post, orphan works make up about 50% to 70% of books published after 1923.

Some publishing industry experts predict that Congress may soon tackle the issue of orphan works. Indeed, in 2008 Congress considered, but did not pass, a bill that would allow the digitization of orphan works if the people or organizations scanning the books could prove they had attempted to locate the rights holder.

Even further complicating legal questions surrounding Google's project is the diverse nature of the collection. Books involved range from fiction titles—which hold some of the highest levels of copyright protection—to nonfiction reference books such as encyclopedias, which, according to the Congressional Research Service, "are afforded the 'thinnest' copyright protection."

In 2008, after years of negotiations, Google reached a settlement with the authors and publishing groups that had sued over the book project in 2005. The settlement provided that Google would compensate rights holders $125 million for use of their works and in return would enjoy a non-exclusive license to digitize and publish snippets of books published before 2009. Under the settlement, Google would have also had a continuous license to scan and sell orphan works. The settlement would have required Google to establish a nonprofit "Book Rights Registry," which would seek to locate rights holders for orphan works and, according to Google, create a "financial incentive" for authors "to come forward" and claim rights over their works.
Also as part of the settlement, Google agreed to exclude most foreign works from its digital library at least temporarily, partially because of protests from European governments, particularly in France and Germany, over copyright issues. Although the agreement would lower the number of books in Google's project by more than half, Google would retain books published in Australia, Canada and Great Britain. Furthermore, the settlement restricted access to the book search to U.S. residents.

The settlement was met with a wide variety of reactions from organizations and individuals that would be affected by it. The American Library Association (ALA), the oldest and largest organization in the world representing libraries, filed a legal brief generally supporting the settlement but expressed some concerns that, if Google obtained a monopoly over the digital book marketplace, it could set subscription rates too high for most libraries to afford. The association thus called for the "rigorous oversight" by the court for the implementation of the settlement.

Many copyright holders, including the family of the late American author John Steinbeck, reversed earlier opposition to Google's project after the terms of the settlement were negotiated. Others, however, objected to the agreement. Corporate competitors of Google, such as Amazon and Microsoft, protested the settlement, saying it would give Google a monopoly over online book searches that would likely violate antitrust laws. In September 2009, authors, publishers and other interests both for and against the settlement filed about 500 court documents.

The Justice Department also opposed the settlement. In April 2009, the department opened an inquiry into the book project because of concerns that the settlement would violate antitrust laws by essentially granting Google a monopoly. In particular, some observers interpreted one clause in the settlement as forbidding authors and publishers involved in the lawsuit from negotiating digital library deals with other companies and giving those companies better terms than they gave Google in the class-action settlement. In response to antitrust concerns, Google eventually dropped the clause from the agreement.

Judges, however, must approve settlements in class action lawsuits before they can be adopted. In March 2011, Judge Chin rejected the settlement agreement, arguing that it would violate antitrust laws by giving Google a "de facto monopoly" in online book searches. Although he acknowledged that a large-scale digital library "would benefit many," he said the settlement was "not fair, adequate and reasonable." Chin also urged Google to switch from an opt-out system to an opt-in model for authors and publishers. Many observers said, however, that Google would reject the opt-in model because it would undermine its main interest in the settlement—the ability to facilitate the inclusion of millions of books in its digital library without further negotiation—and would likely result in the exclusion of millions of orphan works from the collection. [See Excerpts from Court Opinion on Google Books Lawsuit (sidebar)]
Google's Project Tramples Copyright, Antitrust Laws, Critics Say

Where the potential for censorship exists and access to information is concerned, the government must be heavily involved in protecting both the freedom of information and the consumer, critics say. Brewster Kahle, the founder of the Internet Archive, a nonprofit library in San Francisco, writes that an agreement such as the proposed settlement between Google and authors could provide "a new and unsettling form of media consolidation" that would "produce not one but two court-sanctioned monopolies. Google will have permission to bring under its sole control information that has been accessible through public institutions for centuries." Critics note that Google has been willing to censor material before, particularly with its operations in China. [See U.S. Internet Companies in China]

Indeed, critics say, the Google book project threatens to restrict the distribution of information. Kahle writes, "In essence, Google will be privatizing our libraries." He continues:

*Broad access is the greatest promise of our digital age. Giving control over such access to one company, no matter how clever or popular, is a danger to principles we hold dear: free speech, open access to knowledge and universal education. Throughout history, those principles have been realized in libraries.*

Google's book project also poses the danger of monopolizing the book market, critics say. Pamela Samuelson, a copyright expert and law professor at the University of California at Berkeley, contends that a Google lawsuit settlement "is not really a settlement of a dispute over whether scanning books to index them is fair use. It is a major restructuring of the book industry's future without meaningful government oversight."

Some critics of the Google book project argue that the creation of a massive digital library should be a public endeavor, not subject to the control of a private corporation. Robert Darnton, a history professor at Harvard
University and director of the Harvard University Library, writes, "[W]e should not abandon Google's dream of making all the books in the world available to everyone. Instead, we should build a digital public library, which would provide these digital copies free of charge to readers."

A publicly funded project could easily mirror the scale of Google's proposed project, critics argue. Kahle writes, "For the cost of 60 miles of highway, we can have a 10 million–book digital library…. Through a simple Web search, a student researching the life of John F. Kennedy should be able to find books from many libraries, and many booksellers—and not be limited to one private library whose titles are available for a fee, controlled by a corporation that can dictate what we are allowed to read." A public program whose goal is to encourage reading, rather than a private corporation seeking to make a profit, should be in charge of a digital library, critics say.

Keeping such a project out of the hands of a single private corporation, critics say, is necessary to protect the freedom of information. Darnton lauded Judge Chin's March 2011 decision to reject Google's settlement agreement with litigants, writing, "This decision is a victory for the public good, preventing one company from monopolizing access to our common cultural heritage."

The settlement that would have allowed Google's project to continue was insufficient in protecting the rights of both writers and readers, critics argue. Darnton writes, "[T]he settlement didn't do what settlements are supposed to do, like correct an alleged infringement of copyright, or provide damages for past incidents; instead it seemed to determine the way the digital world of books would evolve in the future."

The question of orphan books in particular, critics argue, is a public matter, not one that is relevant only to the participants in one lawsuit. Samuelson told the New York Times, "Even though it is efficient for Google to make all the books available, the orphan works and unclaimed books problem should be addressed by Congress, not by the private settlement of a lawsuit."

Indeed, any judicial settlement of Google's legal woes related to the book project would be insufficient as long as it does not represent all authors and publishers, critics maintain. Samuelson writes, "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?"

A Google-dominated digital library would stunt competition and the growth of e-publishing, critics argue. A brief filed by the Open Book Alliance, an association of writers and publishers formed in protest of the Google settlement, states, "Antitrust laws guarantee consumers the benefit of real competition in the real marketplace—not Google's idea of what competition should produce." Furthermore, the brief argues that, "left unfettered," the new e-publishing industry "offers endless opportunities for authors and small publishers, and a technological bonanza for consumers. But forcing all the new vendors to depend on a single source, Google,…is the surest way to retard what are otherwise boundless prospects."

Critics say that Google's ambition to organize the world's information has made it short-sighted and blind to the possible dangers of such a large-scale, privately operated digital library. Lawrence Lessig, a law professor and director of the Edmond J. Safra Center for Ethics at Harvard University, told the New York Times, "I've seen these big powerful companies filled with people who drank the Kool-Aid. I really get the sense in which these people feel they are doing good. But I am always surprised by their failure to recognize how they will be perceived outside."

**Google's Digital Library Will Allow Researchers Unprecedented Access to Information, Supporters Say**

By making millions of books available to anyone with an Internet connection, rather than just those few who live near the world's largest libraries, the Google book project will level the playing field for students, readers and researchers from all over the world, supporters say. David Balto, a senior fellow at the Center for American Progress, a think tank based in Washington, D.C., writes in the Huffington Post, "[T]he Books project will serve as a
democratizing force across socioeconomic and geographic barriers. Scholars and historians at the smallest schools in the remote corners of this country will obtain the same access to knowledge as those at large well-funded universities in our biggest cities." Furthermore, supporters note, the book project will make many more texts available for the visually impaired, who would be able to read through either screen enlargement or Braille display technology.

For lack of a digital repository, many books and texts are lost to the reader or researcher, supporters note. Google's all-encompassing book project, according to supporters, could remedy this flaw in the publishing cycle. Sergey Brin, a computer scientist who co-founded Google along with Larry Page, writes:

*Books written after 1923 quickly disappear into a literary black hole. With rare exceptions, one can buy them only for a small number of years they are in print. After that, they are found only in a vanishing number of libraries and used book stores. As the years pass, contracts get lost and forgotten, authors and publishers disappear, the rights holders become impossible to track down.*

Furthermore, supporters say the public library system is no substitute for a more secure, larger digital library. Brin writes, "Inevitably, the few remaining copies of [out of print] books are left to deteriorate slowly or are lost to fires, floods and other disasters." Supporters often cite Ancient Greece's famous Alexandria library, which burned three times, and the U.S. Library of Congress, which lost two-thirds of its collection in an 1851 fire, as examples of renowned libraries, now lost to posterity, that could have benefited from digitization, had the technology been available.

Indeed, supporters say that the rejection of the Google settlement denies readers access to millions of books that are out of print or hard to find. James Grimmelmann, a professor at New York Law School in New York City, said, "[T]he settlement could have meant orphan books being made available again. This is basically going back to status quo, and orphan books won't be available."

Although public institutions could have attempted a government-backed large-scale digitization of libraries, supporters note, they have not done so. Google has proved to be the one organization with the ambition, drive and resources to tackle the project, supporters say. Brin writes, "I wish there were a hundred services" like Google's digital library, which

*would have spared Google a tremendous amount of effort. But despite a number of important digitization efforts to date...none have been at a comparable scale, simply because no one else has chosen to invest the requisite resources. At least one such service will have to exist if there are ever to be one hundred.*

Indeed, the government will not likely fund or back any large-scale digital library soon, supporters of Google's project contend. Slate contributor Tim Wu writes, "Some people think it should be the government, not Google, that creates online libraries: Don't hold your breath." Wu maintains that the Google book project could mirror public utility models (like the sewer system), which are generally run by private companies granted de facto monopolies by the government. He argues, "The Anglo-American tradition, in fact, is to put a private company in charge of such public callings, especially those that require large investments. And in the big picture, that is what's really going on here: the creation of an unusual kind of public utility to provide better access to old and unpopular books."

Furthermore, supporters say, critics should not punish Google for its pioneering spirit. Balto writes, "As Judge Learned Hand instructed half a century ago, the antitrust laws are not intended to punish 'superior skill, insight, and industry.'" Balto notes that when Google launched its Book Search project, the company "had to develop its own scanning technology, negotiate numerous agreements with libraries, and navigate the uncertainty surrounding complex copyright issues. Its ability to do all of these things led to a virtual library that offers an unprecedented level of access to millions of consumers."

Indeed, persecuting Google for its drive, some supporters warn, could result in a lost opportunity to secure millions of texts in a digital archive. Wu writes, "It's unlikely that anyone else will take on a money-losing project to scan
Complaints of a potential monopoly by Google, supporters contend, are overwrought. Technology website CNET writer Larry Downes argues, "The problem, according to most of these complaints, is that the agreement gives too much market power to Google over out-of-print books. How's that again? Out-of-print books, by definition, are those for which there is no market today, nor likely to be one any time in the future." Google, supporters say, is resurrecting a market, notcornering it, and the company's legal compromises are paving the way for other companies to embark on similar projects.

Google should move forward with its book project, supporters argue, despite objections. David Drummond, Google's chief legal adviser, said, "The benefits far outweigh any of these criticisms that are being made, many of which are quite theoretical…. The fact that there are some critics doesn't mean you should be paralyzed and not do something that provides value."

Supporters admit that the Google book deal could pose monopoly and copyright problems in the future but hold that the risks of not pursuing the project outweigh such concerns. An editorial in the Economist argues, "[A]ntitrust concerns could surface over time in this brave new digital world. But the theoretical dangers these pose should be weighed against the very real and substantial benefits that a comprehensive digital library will create…. If the court rejects the deal, much potentially useful information will remain, quite literally, a closed book."

**Future of Google's Book Search Project**

The next step in the Google lawsuit is uncertain, although most observers predict Google will continue to work out a settlement to allow its Book Search project to go forward. The authors and publishers who have litigated against Google are similarly unlikely to drop their lawsuits.

Just as critics of Google's project have urged Congress to tackle copyright questions surrounding orphan works, some observers predict that Google too will look for legislative guidance. According to New York Times journalist Claire Cain Miller, "Google may take the battle from the courtroom to Congress, to promote a law that would make orphan works…widely available."

Another possibility, as Judge Chin has suggested, is changing Google's opt-out framework to a structure in which Google can post works only when the rights holder specifically opts in to the agreement. Google spokespeople, however, have criticized an opt-in framework as unworkable given the size of the project.

*Forbes* commentator Quentin Hardy writes, "The present and future of books is clearly digital. The real issue is how these will relate to the past nature of books, which was one of discrete objects that had effectively finite lives." The publishing industry, experts say, will continue to grapple with problems connected to the seemingly inevitable expansion of the digital book market.

**Discussion Questions**

1) Do you think Google should continue to pursue its goal of scanning and indexing every book in the world? Why or why not?

2) Do you think Google's book project constitutes "fair use" under copyright law? Explain your position.

3) Some observers have urged the U.S. to establish a publicly funded and operated national digital library much like those in other countries, such as Norway. Do you think that is possible? What might be some of the roadblocks to
such a project? What would be the advantages and disadvantages of having a public digital library, versus one built by a private company such as Google?

4) Do you think Google's digital library would help or hurt the publishing industry? Explain your position.

5) Research the Google Book Search lawsuit a bit further. If you were the judge deciding the case, would you approve the settlement between Google and its litigants? Why or why not? Write your decision from the point of view of the judge.

Bibliography


**Additional Sources**

Additional information about digital libraries can be found in the following sources:


**Contact Information**

Information on how to contact organizations that either are mentioned in the discussion of digital libraries or can provide additional information on the subject is listed below:

Google Books  
1600 Amphitheatre Parkway  
Mountain View, Calif. 94043  
Telephone: (650) 253-0000  
Internet: [books.google.com](http://books.google.com)

Internet Archive  
300 Funston Avenue  
San Francisco, Calif. 94118  
Telephone: (415) 561-6767  
Internet: [www.archive.org](http://www.archive.org)

Project Gutenberg Literary Archive Foundation  
809 North 1500 West  
Salt Lake City, Utah 84116  
E-mail: help2010@pglaf.org  
Internet: [www.gutenberg.org](http://www.gutenberg.org)

**Keywords**

*For further information about the ongoing debate over digital libraries, search for the following words and terms in electronic databases and other publications:*

Authors Guild  
Book digitization  
Google Books  
Open Knowledge Commons  
Orphan works