

# Open Access and Books in a Digital World - What Role Should Libraries Play?

Barbara Fister

ALA OITP Annual Retreat, Washington DC, November 19 2008

*"If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of everyone, and the receiver cannot dispossess himself of it. Its peculiar character, too, is that no one possesses the less, because every other possesses the whole of it. He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire, expansible over all space, without lessening their density at any point, and like the air in which we breathe, move, and have our physical being, incapable of confinement or exclusive appropriation." - Thomas Jefferson*

Mitch Freedman asked me to join this conversation so that we could examine open access issues beyond the realm of scholarly communication by considering the different approaches to mass book digitization offered by [Google](#) and by the [Open Content Alliance](#). While I'm no expert on these projects, I have long been interested in those parts of culture that are not focused on scholars talking to other scholars. And I'm interested as a librarian in how our values – and our experience with readers – could inform the future of the book industry, which is desperate to find new readers, yet oddly hostile to sharing books, in a digital world where sharing is how things work.

The dynamics of scholarly publishing and other kinds of publishing are quite different. For the author-scholar, the value of publishing research is in its being read and cited by others. Not only do scholars fulfill their mission to add to what we know about the world, but on a more mercenary level, a publication can enhance their CV, earn promotion, and win grants. The value in their work depends on its being circulated and shared.

The writer who relies on advances and royalties on sales of published texts for income has a different relationship to the market. For the commercial writer, gaining readers is important, but only if it increases sales, because in most cases that's what supports their work (though, to be honest, most writers, like most actors, are supported by day jobs). For these writers, libraries are a dodgy business – even though they purchase books (good), they share them among many readers (bad). In these cases, the writer has to consider a less tangible benefit – the development of a healthy reading culture or, on a more practical level, a reader who may want to purchase the author's next book or who may urge the library to do so. Building a relationship with readers will lead to more reading and so, more sales.

In a digital world, however, the rules are changing. In 1994, [John Perry Barlow](#) argued that old notions of property ownership do not work in the digital world. "In the absence of the old containers, almost everything we think we know about intellectual property is wrong. We're going to have to unlearn it. We're going to have to look at information as though we'd never seen the stuff before." Current law focuses on protecting the container of ideas so it's contents can't be altered or copied, but on the Internet, information and expression spill out of containers. It takes special measures far more draconian than those used in the print world to contain digital information. And those measures often conflict with common sense. An academic library director I know recently wondered how many of us were loaning Kindles. I told him that we didn't, but that I understood the Terms of Service prohibited libraries from loaning a Kindle unless there were no books on it. He said loaning a Kindle without books made no sense and that he hadn't actually read the ToS. It seems somehow emblematic of our times that what makes sense is against the rules.

These rules, of course, affect the analog world too. Something like 70% of books are under copyright but not commercially available. An enormous number of those books are "orphans" – we have to seek permission to use

them, but we don't know whom to ask. Or they may have actually entered the public domain, but we can't be sure, so for all practical purposes they are unavailable. The penalties for making an honest mistake are high. The fine for a single act of innocent infringement can be as high as \$30,000.

Yet there is a persistent dream of creating a digital library of all books, making them available to all. Jorge Luis Borges wrote about it in his 1941 story, "[The Library of Babel](#)." His narrator describes a library so vast it has no exact center and no circumference. It contains an infinite number of books – a prototype of the Internet. "When it was announced that the Library contained all books, the first reaction was unbounded joy . . . There was no personal problem, no world problem, whose eloquent solution did not exist." But before long, "[t]his unbridled hopefulness was succeeded, naturally enough, by a similarly disproportionate depression. The certainty that some bookshelf in some hexagon contained precious books, yet that those precious books were forever out of reach, was almost unbearable." There are both dystopian and utopian ways of thinking about the library of everything. But the popular rhetoric about libraries moving online is usually hopeful, and is often expressed as being a way of liberating books from their imprisonment in libraries.

This is somewhat curious. In an age when the NEA tells us reading is in precipitous decline and Steve Jobs announces nobody reads books, the desire to bring books to the Internet is a widely-shared obsession. Over half a million people – a population characterized by one blogger as "[the OCD literati](#)" - have created [LibraryThing](#) accounts and have cataloged 33 million books, with 42 million tags and half a million reviews, and LibraryThing is only one of several popular social networking sites devoted to sharing books. This week, I saw a television ad for Qwest that seemed to suggest they could deliver every book in the Library of Congress to everyone's desktop, and they had snazzy visuals to illustrate it. In fact, they are simply pointing out that, after pumping up their capacity for the DNC in Denver, they had enough capacity to transmit the content of LC, if their math is correct. They could have used the same math to use Netflix holdings as an example, but the idea of a vast library is somehow much more compelling. Books and libraries hold a special place in our collective imagination, just as the moon does when we quantify things by how many of them, stacked up, could reach it. Reaching the moon is an aspiration that contains more than distance, and libraries are a symbol of something more than mere information. It's telling that, when Jeffrey Toobin wrote about Google's library project he titled the article "[Google's Moon Shot](#)." That's how symbolic this quest is to us – as significant a step forward for humankind as landing on the moon.

So efforts to put libraries on the Internet have been ongoing. [Project Gutenberg](#), powered by volunteer labor, has been creating downloadable low-bandwidth text files of public domain books since 1971. It now has around 25,000 books available with nearly 350 added every month. There is also the ambitious [Million Books Project](#) which has scanned and made freely available far more than a million books, most of which are in the public domain because they are old or are government reports. At first glance, this may not be anyone's dream library, unless perhaps you are a County Extension agent.

The Internet Archive, founded by Brewster Kahle in 1996, became the host for a new initiative, the [Open Content Alliance](#), in 2005. It was largely a response to the much-ballyhooed announcement a year earlier of the Google Books library project – an alternative that would be more open and not driven by a profit motive. Interestingly, when the project was announced to the press, *The New York Times* report had the headline, "[In Challenge to Google, Yahoo Will Scan Books](#)." Though Yahoo was just one of many partners, most of which were non-profit, the Times's business desk immediately framed it as a skirmish in the battle for market share among corporations. In fact, many of the libraries contributing to the OCA joined to avoid the restrictions and the commercialization of the Google project and to offer an alternative.

There are five principles for the OCA: The OCA will provide the greatest possible access and reuse of collections while adhering to copyright law. Contributors can set their own restrictions. The OCA can choose what is included. The OCA will provide item and collection level metadata and will encourage creation of more finding tools. The OCA will reside in multiple archives to ensure preservation and access.

In practice, the archive is contributor-driven. Participating libraries pay the freight. The process is more costly and slower for libraries than Google's process. And its principle of respecting rightsholders' interpretation of copyright gives Google an enormous advantage.

Google began its book project after [Amazon](#) launched an amazing new tool - Search Inside the Book - in Fall, 2003.

Overnight searching for information in books was transformed. This came as a total surprise to copyright holders, whose books were suddenly full text at Amazon. The company had secretly worked out the deal with publishers, who almost always acquire electronic rights when they secure first publication rights. It was mind-blowing. And Google had a serious contender – at least when it came to the contents of current US books. Of course, you could only view five consecutive pages, you couldn't copy and paste or print, and you had to turn over both your credit card and your privacy to use the service. (Amazon promised publishers they would receive data on what we were looking up and which chunks of books were most popular as part of the deal.) But what would have been deeply disturbing to people ten or twenty years ago was now the price you paid for the convenience of search. We paid it every day when we used Google.

Google tried to get its foot in the door with Google Print. They offered publishers a platform for book information including excerpts. Publishers yawned. Amazon already provided that, and they're a bookseller, besides. Frustrated, Google checked their bank account (deep? Check) and took an audacious step. They turned to libraries, notorious for profligate sharing, and offered them a great deal. We'll scan your collections. You'll get a copy. And we'll pay the lawyers. For Google, it was a shrewd piece of public relations. They were able to capitalize on the general belief that books (even if published in 1886) offer higher quality content than Websites. They gained the prestige of co-branding themselves with Harvard, Oxford, and Stanford. And those august names lent credence to the implication that soon Google really would be the source of all knowledge. Their announcement scored the front page of *The New York Times*. Buried in the story was the fact that Harvard and Oxford had opted only to digitize selected public domain works. But other libraries were excited about the potential to call the question on what fair use means for digital collections. I thought that would be the real benefit of the project: a company with deep enough pockets to afford the risk would take this question to the courts. A loss would be a disaster, but a win could open up a new era of possibility. Jeffrey Toobin explored both possibilities in a 2007 *New Yorker* article. He predicted that Google would settle and that would be a blow, because it would mean there would be no competitors in the race to offer the universal library. Google would be the winner, even if someone came along with a better product.

Forward to fall of 2008. Google announces [a settlement](#). And reactions are mixed.

Many commentators are delighted that more content will become visible if the court approves the settlement as is. Not only will the snippets grow to the size of gobbets, but those books that are no longer on the market will be available for sale. This is seen by many as the biggest leap forward: that 70% of books that were under copyright but not commercially available would now be back in circulation.

Some of my friends who are avid readers were delighted that they would be able to obtain long out-of-print volumes and that the authors would get some revenue, unlike a used book sale. Without meaning to rain on their parade, I pointed out that it may take years to work out the deal, that "available" was iffy, since reading it online was not what they had in mind and Google scans of library books are not print quality. And even more to the point, it was unlikely that the popular fiction they were looking for would be on the shelves of the University of Michigan libraries. Buried in the hype of news stories a simple point is missed: Google has not scanned every book ever published. They will not be able to provide access to them all. There are limits to what books will be available and in what manner they will be available.

There have also been concerns about how Google defines "public domain." Mass digitization projects tend to amplify legal restrictions simply because they are massive; they can't be bothered to determine rights book by book. Government documents scanned in the Google library project are restricted from view, even though they are in the public domain. Books published between 1923 and 1976 that may well be in the public domain are treated as if they are under copyright. And Google Books (and its altruistic cousin, the [Hathi Trust](#)) act differently if you're in another country. Peter Suber [has reported](#) that books that are demonstrably in the public domain in all countries of the world are nevertheless restricted "due to copyright limitations." In fact, it's due to mass digitization limitations. No time to sort out the rights, so restrict everything that may pose a problem.

The not-for-profit registry that the settlement describes will allow authors and publishers to profit from sales originating with Google searches and will allow readers access to otherwise unavailable books. It will be more effective adoption agency than that dreadful mess of legislation that was intended to govern orphan works but hasn't yet passed. According to [Georgia Harper](#), there's an efficiency about this.

This isn't the Congressional approach to problem solving (shove the parties into a room and lock the door until they have reached an agreement -- and may the strongest interest obliterate the weaker and we'll call it a compromise in the public interest). This is the publisher's and Google's no nonsense business approach: "Hey, let's just start selling all the books and if there's money to be made, the owners will either show up to claim it, or the money will lie there for 5 years while we give everyone time to wake up and smell the coffee. At the end of 5 years, we'll pretty much know what's orphan and what's not.

But it also reframes the idea of who should benefit when an orphan work generates a sale. In the terms of the registry, the registry keeps the money. This means sale of orphan works benefits authors and publishers, but not the public. It also may invite specious ownership claims that will be difficult to resolve. Some commentators also worry that, since the agreement only speaks of online access to purchased books and to public domain titles, that the current availability of downloadable .pdfs may be discontinued at Google's whim. And then there's an overarching concern: when a single for-profit company provides the access to so much information, have we created a monster? How much power is vested in one not-so-transparent entity?

We can look at this pragmatically, as [Paul Courant](#) of the University of Michigan libraries does. From his perspective, Google has accomplished something that libraries on their own could not and that the Open Content Alliance could not. They digitized seven million books in a very short period of time. They had the deep pockets to take risks that libraries could not afford. They did not provoke an answer that would clarify fair use, but will be able to provide greater access than they could have done with a fair use settlement, where only snippets would remain available. They've streamlined and centralized rights management for the public. And they did it without it costing the libraries who not only keep the books that were digitized, but can keep their digital copies.

[Georgia Harper](#) adds another intriguing potential benefit to the public. Google will not only open a market for books that were not available to the marketplace, Google's manipulation of book prices will be so massive and yet so malleable that they are in a position, as the new middleman between books and consumers, to conduct experiments that could find the balance between totally open access and revenue-generation. She thinks they will learn – and will have the evidence to prove it – that the optimal price for digital copies is zero. Instead of pay per view, open access will stimulate print sales.

This would not surprise writers like [Cory Doctorow](#), who have conducted their own experiments to show that a book freely available on the 'net can nevertheless hit *The New York Times* bestseller list, as [Little Brother](#) recently did. Publishers spend all kinds of money to get attention for their new releases. They send out hundreds of printed advanced reader copies (ARCs) to newspapers, even though book review space is shrinking drastically. To make up for lack of reviews, they have taken to distributing ARCs to bloggers and they even pay for the privilege of having LibraryThing give them away through their Early Reviewer program. So why is it so hard for them to make the leap from giving away lots of printed copies for free to giving them away digitally?

They watched the music industry tank and fear that they are next. Rather than trying to figure out what the music industry did wrong, they worry about pirates. They're afraid that digital copies, once free, will scamper about and reproduce like rabbits. But would that be all bad? When I was looking for websites of popular Scandinavian authors for a recent project, I found that many of them didn't bother to have websites, which seemed odd, since it's considered a standard marketing tool in the US; instead, the first links in a Google search were to torrent sites offering downloads. Yet that doesn't stop these books from selling millions of copies. Paul Coelho saw [pirated copies](#) of his books in multiple languages and decided to [start his own piracy program](#). His publishers weren't happy about it, but they changed their minds when his sales soared. What Georgia Harper is suggesting is that data from Google's pricing experiments might be just what is needed to overcome publisher's fear of open access.

Still, some significant ethical issues remain. We woke up one morning and found out that libraries had colluded in making Google the world greatest aggregator and reseller of book content, creating out of libraries yet another product that libraries will have to subscribe to. In [an interview](#) with the *San Jose Mercury News*, Brewster Kahle said "When Google started out, they pointed people to other people's content. Now they're breaking the model of the Web. They're like the bad old days of AOL, trying to build a walled garden of content that you have to pay to see."

Possibly the best critique of the Google settlement was written four years ago by Rory Litwin, who saw the long-term outcome of the Google library project much more clearly than I did. Shortly after the library project was

announced, in an essay titled "[Google and the Monetization of Libraries](#)" he wrote:

Though they have not announced plans to offer the full text of copyrighted materials on a pay-per-view basis, with fees turned over to copyright owners, it is a technical possibility with the natural force of an economic vacuum in the corporate context. Logically it would seem to be only a matter of time before this mode of access becomes a reality, providing a channel for bypassing both public-interest information policies and the librarian's professional service. . . .

Google is not us. Google is not staffed by librarians, and does not operate according to policies that flow out of long traditions of library practice guaranteeing privacy, equity of access, collective ownership of information, information in context, and personal service. This project, as Larry Page has already put it, is about *monetizing* the holdings of research libraries. It is about commercializing library collections that it has taken centuries to build. It may be the "greatest information conglomerate of all time," but it is not us. We are nowhere in it; we do not control it or even influence it. We may be invited to imagine that it is "us," that it is "a library" or even that it is "Library," and we may be flattered by the attention, but we should take care to remember what librarianship means in contradistinction to commercialized information, to remember the difference between individuals-as-citizens and individuals-as-consumers and to remember that as librarians we are public stewards of the information commons and have an obligation to preserve and protect it. And, to say it one last time, we must not let anyone write off these concerns as "sentimental." They are not; what they are is simply values-driven.

This, for me, raises a lot of questions. What do our values lead to in this era in which expedience trumps values? In which libraries do not have the funds to do what business can do, so we give up things we used to consider sacrosanct? Does reader privacy mean anything in an age when our private lives are the currency of digital commerce? What public good do we provide when an increasing portion of our budgets goes toward purchasing temporary and limited access to walled gardens tended by corporations whose customers are not just libraries, but publishers?

We should be having these conversations. We should become activists, not just on behalf of scholarly communication, but all human expression. Our values may not be entirely incompatible with commerce; I'm convinced that what we know about sharing could actually help the publishing industry reinvent itself. But I'm worried instead that the publishing industry will reinvent libraries, and we'll be left with values that we can't afford to practice.

## **Mass Digitization Projects**

### [Google Books](#)

A digitization project that, when publishers were reluctant to join in, recruited research libraries. The recent settlement of a class action lawsuit has spurred much comment.

### [Hathi Trust](#)

A collaboration of the thirteen universities of the Committee on Institutional Cooperation and the University of California system to establish a repository for these universities to archive and share their digitized collections, including books scanned in the Google library project. Other libraries may join. The goal is to provide persistent, secure storage for digital collections. Those in the public domain will be accessible to all; others will be accessible depending on the wishes of rightsholders. This recently-launched project currently holds almost two million volumes, of which 500,000 are full-text searchable. Though much of the content is copies of books scanned by Google, the intention is to add digital collections not in Google Books.

### [Open Content Alliance](#)

A not-for-profit collaborative to digitize books and make them freely available. About a million books are currently

included. Related to the Open Library project which hopes to include a record of every book ever published and currently lists about 20 million records.

#### [Project Gutenberg](#)

An example of the early potential of the Internet to distribute out-of-copyright books on a volunteer-labor basis.

#### [Universal Digital Library](#) (Million Books Project)

An international effort based at Carnegie Mellon University to provide worldwide free access to books. Its goal: "to capture all books in digital format." The collection contains over 1.5 million mostly out-of-copyright books and many government documents, with an emphasis on agriculture. They plan to host ten million books in ten years.

#### [EEBO - Early English Books Online](#)

An example of a book digitization project that is a "walled garden" - libraries must pay a substantial fee and annual access costs and must limit access to their students, faculty, and staff. Thus the libraries' investment comes with limits that physical books do not have.

#### **Other Open Access Book Projects**

##### [OAPEN: Open Access Publishing in European Networks](#)

An open access publishing project for humanities and social sciences involving six European university presses but intended to grow as it develops an e-publishing platform. The project is meant to establish a more economic alternative to traditional publishing for small-audience not-for-profit presses and to spread OA principles from STM publishing to the humanities and social sciences.

#### **OA is Good for Business - Really!**

Cory Doctorow makes his books available online for free using a Creative Commons License. Free access does not interfere with his books hitting bestseller lists. See [Peter Suber's coverage](#) of *Little Brother* or browse [Content](#), his most recent collection of essays on copyright and intellectual property.

Paul Coehlo [decided to round up](#) torrent sites offering pirated copies of his books in (often unauthorized) translations and found his book sales soared. More on the [Guardian Books Blog](#).

#### [Bloomsbury Academic](#)

A new imprint that combine a print list with Creative Commons-licensed copies online.

#### [Rice University Press](#)

A once-defunct press is revived as an open access project with print on demand options based on the [Connexions](#) collaborative open access publishing platform.

#### [Yale University Press Books Unbound](#)

A selection of Yale UP books made available under Creative Commons licenses, including Yochai Benkler's [Wealth of Networks](#) and Jonathan Zittrain's [The Future of the Internet and How to Stop It](#). Readers can add annotations and comments using Intitute for the Future of the Book's Commentpress software.

#### **Dipping a Toe in the Open Access Waters**

##### [Romance Studies](#) (Penn State Press)

A university press, in collaboration with its library and an academic department, revived a defunct series in open access mode because it makes more economic sense than print does for this limited-audience material.

In the past year, several commercial publishers have been releasing books in electronic form for a limited time as a promotion. Are they testing the waters for more extensive OA?

### **About the Google Books library project and settlement**

#### [Google Book Search Copyright Settlement](#)

The official site with the text of the settlement agreement that still awaits court approval.

Band, Jonathan. [A Guide for the Perplexed: Libraries and the Google Library Project Settlement](#). ALA & ARL, 11/13/08

A handy guide to the provisions of the settlement that relate particularly to libraries.

Grimmelman, James. "[Principles and Recommendations for the Google Books Settlement](#)." Laboratorium, 11/8/08  
A set of recommendations to the court for changes that would protect the public interest.

Harper, Georgia - "[Google Book Search and Orphan Works](#)." Collectanea, 11/1/08

Harper, Georgia - "[Settlement Controlled Pricing and Tests on Effects of Openness](#)." Collectanea, 11/8/08

The settlement allows Google to play with pricing that will provide the optimal balance for publishers and authors between openness and profitability. Harper predicts this research could establish an informed basis for practices that are now only guesswork.

Lessig, Lawrence. "[On the Google Books Agreement](#)." Lessig 2.0, 10/29/08

Sees much that is good in the agreement, but points out some fuzzy areas where much will depend on its execution.

Litwin, Rory. "[On Google's Monetization of Libraries](#)." Library Juice 7.62, 12/17/04

An important essay on the real costs of libraries' collaboration with Google and a prediction of what would come with the settlement.

"[Let's Not Settle for This Settlement](#)." Open Content Alliance Blog, 11/5/08

"At its heart, the settlement agreement grants Google an effective monopoly on an entirely new commercial model for accessing books. It re-conceives reading as a billable event. This reading event is therefore controllable and trackable. It also forces libraries into financing a vending service that requires they perpetually buy back what they have already paid for over many years of careful collection."

Murray, Peter E. "[Google Book Search Settlement: Reviewing the Notice of Settlement](#)." Disruptive Library Technology Jester, 10/29/08

An analysis of the settlement with libraries in mind.

Sanfilippo, Tony. "[The Google Thing](#)." PSU Press Blog 10/29/08

A perspective from a university press. In the short run, the settlement looks good. "But in the long run, we may have just been forced into an arranged marriage with a groom that clearly has a different set of values than we have. I guess we'll just have to wait and see which of Google's objectives will win at the end of the day—the one that answers to its evil-less mission, or the one that answers to its shareholders." Points out that aggregators who were moving into the book market are likely to be the losers, since Google, with over 7 million books digitized, will be a hard act to follow.

Toobin, Jeffrey. "[Google's Moon Shot: The Quest for the Universal Library](#)." *The New Yorker* 2/5/07.

Predicted the settlement and the fact that it would leave Google in a strong position - which would be bad news for competition and for the public.

van Lohmann, Fred. "[Google Book Search Settlement: A Reader's Guide](#)." Deeplinks Blog (Electronic Frontier Foundation), 10/31/08.

A legal analysis that examines its impact on innovation, competition, access, privacy, fair use, and the public domain.

### **About Book Digitization Projects Generally**

Grafton, Anthony. "[Future Reading: Digitization and its Discontents](#)." *The New Yorker* 11/5/07

A somewhat nostalgic historical perspective, skeptical of digitization as a universal library - or as a replacement for libraries with printed books.

Harper, Georgia. "[Mass Digitization and Copyright Law, Policy and Practice](#)." 5/08

An excellent overview, with a thorough list of sources.

### **About openness versus locked-up "intellectual property"**

Barlow, John Perry - "[The Economy of Ideas](#)." *Wired* 3/94

A classic article that delineates the ways digital expression evade the concept of "property." Barlow says "the increasing difficulty of enforcing existing copyright and patent laws is already placing in peril the ultimate source of intellectual property - the free exchange of ideas." His vision: information is not static but interactive, that it can't be possessed, that it is a relationship rather than an object. "Information economics, in the absence of objects, will be based more on relationship than possession."

Greco, Albert and Robert M. Wharton. "[Should University Presses Adopt an Open Access \[Electronic Publishing\] Business Model for All of Their Scholarly Books?](#)" Proceedings of the 12th International Conference on Electronic Publishing held in Toronto, Canada 25-27 June 2008.

Examines UP sales data from 200-2007 and concludes open access makes economic sense. Competition from other publishers makes it hard for UPs to find their niche. Open access better matches their mission and would be less costly (without sacrificing editorial quality).

Lessig, Laurence. [Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity](#). Penguin, 2004.

A groundbreaking book on how media has distorted copyright. See also his recent publication [Remix](#), which argues that Read Only culture clashes with Read/Write culture, that redefining ownership will make economic sense, while defining what most young people do as they engage with culture as criminal behavior is creating a generation of so-called outlaws.



This work by Barbara Fister is licensed under a [Creative Commons Attribution-Share Alike 3.0 United States License](#).