

nybooks.com

Google & the Future of Books

by ROBERT DARNTON • FEB. 12, 2009

The newly restored reference room of the library at the American Academy in Rome

How can we navigate through the information landscape that is only beginning to come into view? The question is more urgent than ever following the recent settlement between Google and the authors and publishers who were suing it for alleged breach of copyright. For the last four years, Google has been

digitizing millions of books, including many covered by copyright, from the collections of major research libraries, and making the texts searchable online. The authors and publishers objected that digitizing constituted a violation of their copyrights. After lengthy negotiations, the plaintiffs and Google agreed on a settlement, which will have a profound effect on the way books reach readers for the foreseeable future. What will that future be?

No one knows, because the settlement is so complex that it is difficult to perceive the legal and economic contours in the new lay of the land. But those of us who are responsible for research libraries have a clear view of a common goal: we want to open up our collections and make them available to readers everywhere. How to get there? The only workable tactic may be vigilance: see as far ahead as you can; and while you keep your eye on the road, remember to look in the rearview mirror.

When I look backward, I fix my gaze on the eighteenth century, the Enlightenment, its faith in the power of knowledge, and the world of ideas in which it operated—what the enlightened referred to as the Republic of Letters.

The eighteenth century imagined the Republic of Letters as a realm with no police, no boundaries, and no inequalities other than those determined by talent. Anyone could join it by exercising the two main attributes of citizenship, writing and reading. Writers formulated ideas, and readers judged them. Thanks to the power of the printed word, the judgments spread in widening circles, and the strongest arguments won.

The word also spread by written letters, for the eighteenth century was a great era of epistolary exchange. Read through the correspondence of Voltaire, Rousseau, Franklin, and Jefferson—each filling about fifty volumes—and you can watch the Republic of Letters in operation. All four writers debated all the issues of their day in a steady stream of letters, which crisscrossed Europe and America in a transatlantic information

network.

I especially enjoy the exchange of letters between Jefferson and Madison. They discussed everything, notably the American Constitution, which Madison was helping to write in Philadelphia while Jefferson was representing the new republic in Paris. They often wrote about books, for Jefferson loved to haunt the bookshops in the capital of the Republic of Letters, and he frequently bought books for his friend. The purchases included Diderot's *Encyclopédie*, which Jefferson thought that he had got at a bargain price, although he had mistaken a reprint for a first edition.

Two future presidents discussing books through the information network of the Enlightenment—it's a stirring sight. But before this picture of the past fogs over with sentiment, I should add that the Republic of Letters was democratic only in principle. In practice, it was dominated by the wellborn and the rich. Far from being able to live from their pens, most writers had to court patrons, solicit sinecures, lobby for appointments to state-controlled journals, dodge censors, and wangle their way into salons and academies, where reputations were made. While suffering indignities at the hands of their social superiors, they turned on one another. The quarrel between Voltaire and Rousseau illustrates their temper. After reading Rousseau's *Discourse on the Origins of Inequality* in 1755, Voltaire wrote to him, "I have received, Monsieur, your new book against the human race.... It makes one desire to go down on all fours." Five years later, Rousseau wrote to Voltaire. "Monsieur,...I hate you."

The personal conflicts were compounded by social distinctions. Far from functioning like an egalitarian agora, the Republic of Letters suffered from the same disease that ate through all societies in the eighteenth century: privilege. Privileges were not limited to aristocrats. In France, they applied to everything in the world of letters, including printing and the book trade, which were dominated by exclusive guilds, and the books themselves, which could not appear legally without a royal privilege and a censor's approbation, printed in full in their text.

One way to understand this system is to draw on the sociology of knowledge, notably Pierre Bourdieu's notion of literature as a power field composed of contending positions within the rules of a game that itself is subordinate to the dominating forces of society at large. But one needn't subscribe to Bourdieu's school of sociology in order to acknowledge the connections between literature and power. Seen from the perspective of the players, the realities of literary life contradicted the lofty ideals of the Enlightenment. Despite its principles, the Republic of Letters, as it actually operated, was a closed world, inaccessible to the underprivileged. Yet I want to invoke the Enlightenment in an argument for openness in general and for open access in particular.

If we turn from the eighteenth century to the present, do we see a similar contradiction between principle and practice—right here in the world of research libraries? One of my colleagues is a quiet, diminutive lady, who might call up the notion of Marion the Librarian. When she meets people at parties and identifies

herself, they sometimes say condescendingly, “A librarian, how nice. Tell me, what is it like to be a librarian?” She replies, “Essentially, it is all about money and power.”

We are back with Pierre Bourdieu. Yet most of us would subscribe to the principles inscribed in prominent places in our public libraries. “Free To All,” it says above the main entrance to the Boston Public Library; and in the words of Thomas Jefferson, carved in gold letters on the wall of the Trustees’ Room of the New York Public Library: “I look to the diffusion of light and education as the resource most to be relied on for ameliorating the condition promoting the virtue and advancing the happiness of man.” We are back with the Enlightenment.

Our republic was founded on faith in the central principle of the eighteenth-century Republic of Letters: the diffusion of light. For Jefferson, enlightenment took place by means of writers and readers, books and libraries—especially libraries, at Monticello, the University of Virginia, and the Library of Congress. This faith is embodied in the United States Constitution. Article 1, Section 8, establishes copyright and patents “for limited times” only and subject to the higher purpose of promoting “the progress of science and useful arts.” The Founding Fathers acknowledged authors’ rights to a fair return on their intellectual labor, but they put public welfare before private profit.

How to calculate the relative importance of those two values? As the authors of the Constitution knew, copyright was created in Great Britain by the Statute of Anne in 1710 for the purpose of curbing the monopolistic practices of the London Stationers’ Company and also, as its title proclaimed, “for the encouragement of learning.” At that time, Parliament set the length of copyright at fourteen years, renewable only once. The Stationers attempted to defend their monopoly of publishing and the book trade by arguing for perpetual copyright in a long series of court cases. But they lost in the definitive ruling of *Donaldson v. Becket* in 1774.

When the Americans gathered to draft a constitution thirteen years later, they generally favored the view that had predominated in Britain. Twenty-eight years seemed long enough to protect the interests of authors and publishers. Beyond that limit, the interest of the public should prevail. In 1790, the first copyright act—also dedicated to “the encouragement of learning”—followed British practice by adopting a limit of fourteen years renewable for another fourteen.

How long does copyright extend today? According to the Sonny Bono Copyright Term Extension Act of 1998 (also known as “the Mickey Mouse Protection Act,” because Mickey was about to fall into the public domain), it lasts as long as the life of the author plus seventy years. In practice, that normally would mean more than a century. Most books published in the twentieth century have not yet entered the public domain. When it comes to digitization, access to our cultural heritage generally ends on January 1, 1923, the date from which great numbers of books are subject to copyright laws. It will remain there—unless private interests take over the digitizing, package it for consumers, tie the packages up by means of legal deals, and sell them for the

profit of the shareholders. As things stand now, for example, Sinclair Lewis's *Babbitt*, published in 1922, is in the public domain, whereas Lewis's *Elmer Gantry*, published in 1927, will not enter the public domain until 2022.¹

To descend from the high principles of the Founding Fathers to the practices of the cultural industries today is to leave the realm of Enlightenment for the hurly-burly of corporate capitalism. If we turned the sociology of knowledge onto the present—as Bourdieu himself did—we would see that we live in a world designed by Mickey Mouse, red in tooth and claw.

Does this kind of reality check make the principles of Enlightenment look like a historical fantasy? Let's reconsider the history. As the Enlightenment faded in the early nineteenth century, professionalization set in. You can follow the process by comparing the *Encyclopédie* of Diderot, which organized knowledge into an organic whole dominated by the faculty of reason, with its successor from the end of the eighteenth century, the *Encyclopédie méthodique*, which divided knowledge into fields that we can recognize today: chemistry, physics, history, mathematics, and the rest. In the nineteenth century, those fields turned into professions, certified by Ph.D.s and guarded by professional associations. They metamorphosed into departments of universities, and by the twentieth century they had left their mark on campuses—chemistry housed in this building, physics in that one, history here, mathematics there, and at the center of it all, a library, usually designed to look like a temple of learning.

Along the way, professional journals sprouted throughout the fields, subfields, and sub-subfields. The learned societies produced them, and the libraries bought them. This system worked well for about a hundred years. Then commercial publishers discovered that they could make a fortune by selling subscriptions to the journals. Once a university library subscribed, the students and professors came to expect an uninterrupted flow of issues. The price could be ratcheted up without causing cancellations, because the libraries paid for the subscriptions and the professors did not. Best of all, the professors provided free or nearly free labor. They wrote the articles, refereed submissions, and served on editorial boards, partly to spread knowledge in the Enlightenment fashion, but mainly to advance their own careers.

The result stands out on the acquisitions budget of every research library: the *Journal of Comparative Neurology* now costs \$25,910 for a year's subscription; *Tetrahedron* costs \$17,969 (or \$39,739, if bundled with related publications as a *Tetrahedron* package); the average price of a chemistry journal is \$3,490; and the ripple effects have damaged intellectual life throughout the world of learning. Owing to the skyrocketing cost of serials, libraries that used to spend 50 percent of their acquisitions budget on monographs now spend 25 percent or less. University presses, which depend on sales to libraries, cannot cover their costs by publishing monographs. And young scholars who depend on publishing to advance their careers are now in danger of perishing.



The book-scanning project at the British Library, London, which received funding from Microsoft until that company discontinued its book digitization program last May

Fortunately, this picture of the hard facts of life in the world of learning is already going out of date. Biologists, chemists, and physicists no longer live in separate worlds; nor do historians, anthropologists, and literary scholars. The old map of the campus no longer corresponds to the activities of the professors and students. It is being redrawn everywhere, and in many places the interdisciplinary designs are turning into structures. The library remains at the heart of things, but it pumps nutrition throughout the university, and often to the farthest reaches of cyberspace, by means of electronic networks.

The eighteenth-century Republic of Letters had been transformed into a professional Republic of Learning, and it is now open to amateurs—amateurs in the best sense of the word, lovers of learning among the general citizenry. Openness is operating everywhere, thanks to “open access” repositories of digitized articles available free of charge, the Open Content Alliance, the Open Knowledge Commons, OpenCourseWare, the Internet Archive, and openly amateur enterprises like Wikipedia. The democratization of knowledge now seems to be at our fingertips. We can make the Enlightenment ideal come to life in reality.

At this point, you may suspect that I have swung from one American genre, the jeremiad, to another, utopian enthusiasm. It might be possible, I suppose, for the two to work together as a dialectic, were it not for the danger of commercialization. When businesses like Google look at libraries, they do not merely see temples of learning. They see potential assets or what they call “content,” ready to be mined. Built up over centuries at an enormous expenditure of money and labor, library collections can be digitized en masse at relatively little cost—millions of dollars, certainly, but little compared to the investment that went into them.

Libraries exist to promote a public good: “the encouragement of learning,” learning “Free To All.” Businesses exist in order to make money for their shareholders—and a good thing, too, for the public good depends on a profitable economy. Yet if we permit the commercialization of the content of our libraries, there is no getting around a fundamental contradiction. To digitize collections and sell the product in ways that fail to guarantee wide access would be to repeat the mistake that was made when publishers exploited the market for scholarly journals, but on a much greater scale, for it would turn the Internet into an instrument for

privatizing knowledge that belongs in the public sphere. No invisible hand would intervene to correct the imbalance between the private and the public welfare. Only the public can do that, but who speaks for the public? Not the legislators of the Mickey Mouse Protection Act.

You cannot legislate Enlightenment, but you can set rules of the game to protect the public interest. Libraries represent the public good. They are not businesses, but they must cover their costs. They need a business plan. Think of the old motto of Con Edison when it had to tear up New York's streets in order to get at the infrastructure beneath them: "Dig we must." Libraries say, "Digitize we must." But not on any terms. We must do it in the interest of the public, and that means holding the digitizers responsible to the citizenry.

It would be naive to identify the Internet with the Enlightenment. It has the potential to diffuse knowledge beyond anything imagined by Jefferson; but while it was being constructed, link by hyperlink, commercial interests did not sit idly on the sidelines. They want to control the game, to take it over, to own it. They compete among themselves, of course, but so ferociously that they kill each other off. Their struggle for survival is leading toward an oligopoly; and whoever may win, the victory could mean a defeat for the public good.

Don't get me wrong. I know that businesses must be responsible to shareholders. I believe that authors are entitled to payment for their creative labor and that publishers deserve to make money from the value they add to the texts supplied by authors. I admire the wizardry of hardware, software, search engines, digitization, and algorithmic relevance ranking. I acknowledge the importance of copyright, although I think that Congress got it better in 1790 than in 1998.

But we, too, cannot sit on the sidelines, as if the market forces can be trusted to operate for the public good. We need to get engaged, to mix it up, and to win back the public's rightful domain. When I say "we," I mean we the people, we who created the Constitution and who should make the Enlightenment principles behind it inform the everyday realities of the information society. Yes, we must digitize. But more important, we must democratize. We must open access to our cultural heritage. How? By rewriting the rules of the game, by subordinating private interests to the public good, and by taking inspiration from the early republic in order to create a Digital Republic of Learning.

What provoked these jeremianic- utopian reflections? Google. Four years ago, Google began digitizing books from research libraries, providing full-text searching and making books in the public domain available on the Internet at no cost to the viewer. For example, it is now possible for anyone, anywhere to view and download a digital copy of the 1871 first edition of *Middlemarch* that is in the collection of the Bodleian Library at Oxford. Everyone profited, including Google, which collected revenue from some discreet advertising attached to the service, Google Book Search. Google also digitized an ever-increasing number of library books that were protected by copyright in order to provide search services that displayed small snippets of the text. In September and October 2005, a group of authors and publishers brought a class action suit

against Google, alleging violation of copyright. Last October 28, after lengthy negotiations, the opposing parties announced agreement on a settlement, which is subject to approval by the US District Court for the Southern District of New York.²

The settlement creates an enterprise known as the Book Rights Registry to represent the interests of the copyright holders. Google will sell access to a gigantic data bank composed primarily of copyrighted, out-of-print books digitized from the research libraries. Colleges, universities, and other organizations will be able to subscribe by paying for an “institutional license” providing access to the data bank. A “public access license” will make this material available to public libraries, where Google will provide free viewing of the digitized books on one computer terminal. And individuals also will be able to access and print out digitized versions of the books by purchasing a “consumer license” from Google, which will cooperate with the registry for the distribution of all the revenue to copyright holders. Google will retain 37 percent, and the registry will distribute 63 percent among the rightsholders.

Meanwhile, Google will continue to make books in the public domain available for users to read, download, and print, free of charge. Of the seven million books that Google reportedly had digitized by November 2008, one million are works in the public domain; one million are in copyright and in print; and five million are in copyright but out of print. It is this last category that will furnish the bulk of the books to be made available through the institutional license.

Many of the in-copyright and in-print books will not be available in the data bank unless the copyright owners opt to include them. They will continue to be sold in the normal fashion as printed books and also could be marketed to individual customers as digitized copies, accessible through the consumer license for downloading and reading, perhaps eventually on e-book readers such as Amazon’s Kindle.

After reading the settlement and letting its terms sink in—no easy task, as it runs to 134 pages and 15 appendices of legalese—one is likely to be dumbfounded: here is a proposal that could result in the world’s largest library. It would, to be sure, be a digital library, but it could dwarf the Library of Congress and all the national libraries of Europe. Moreover, in pursuing the terms of the settlement with the authors and publishers, Google could also become the world’s largest book business—not a chain of stores but an electronic supply service that could out-Amazon Amazon.

An enterprise on such a scale is bound to elicit reactions of the two kinds that I have been discussing: on the one hand, utopian enthusiasm; on the other, jeremiads about the danger of concentrating power to control access to information.

Who could not be moved by the prospect of bringing virtually all the books from America’s greatest research libraries within the reach of all Americans, and perhaps eventually to everyone in the world with access to the Internet? Not only will Google’s technological wizardry bring books to readers, it will also open up

extraordinary opportunities for research, a whole gamut of possibilities from straightforward word searches to complex text mining. Under certain conditions, the participating libraries will be able to use the digitized copies of their books to create replacements for books that have been damaged or lost. Google will engineer the texts in ways to help readers with disabilities.

Unfortunately, Google's commitment to provide free access to its database on one terminal in every public library is hedged with restrictions: readers will not be able to print out any copyrighted text without paying a fee to the copyright holders (though Google has offered to pay them at the outset); and a single terminal will hardly satisfy the demand in large libraries. But Google's generosity will be a boon to the small-town, Carnegie-library readers, who will have access to more books than are currently available in the New York Public Library. Google can make the Enlightenment dream come true.

But will it? The eighteenth-century philosophers saw monopoly as a main obstacle to the diffusion of knowledge—not merely monopolies in general, which stifled trade according to Adam Smith and the Physiocrats, but specific monopolies such as the Stationers' Company in London and the booksellers' guild in Paris, which choked off free trade in books.

Google is not a guild, and it did not set out to create a monopoly. On the contrary, it has pursued a laudable goal: promoting access to information. But the class action character of the settlement makes Google invulnerable to competition. Most book authors and publishers who own US copyrights are automatically covered by the settlement. They can opt out of it; but whatever they do, no new digitizing enterprise can get off the ground without winning their assent one by one, a practical impossibility, or without becoming mired down in another class action suit. If approved by the court—a process that could take as much as two years—the settlement will give Google control over the digitizing of virtually all books covered by copyright in the United States.

Original URL:

<http://www.nybooks.com/articles/archives/2009/feb/12/google-the-future-of-books/>