

Font Size: [A](#) [A](#) [A](#)

The New York Review of Books

Google's Loss: The Public's Gain

APRIL 28, 2011

Robert Darnton



Musée du Louvre, Paris

Jean-Baptiste-Camille Corot: La Madeleine lisant, circa 1854; from the Oskar Reinhart Collection's exhibition 'The Secret Cabinet: A Reader in Context,' reviewed by Willibald Sauerländer in this issue

It is too early to do a postmortem on Google's attempt to digitize and sell millions of books, despite the decision by Judge Denny Chin on March 23 to reject the agreement that seemed to make Google's project possible. Google Book Search may rise from the ashes, reincarnated in some new settlement with the authors and publishers who had taken Google to court for alleged infringement of their copyrights. But this is a good time to take a backward look at the ground covered by Google since it first set out to provide access to all the books in the world. What went wrong?

In the forty-eight-page opinion that accompanied his decision, Judge Chin indicated some of the wrong turns and paths not taken. His reasoning ran through each stage in the evolution of the enterprise:

- 2004: Google started digitizing books from research libraries and displaying snippets of them for online searches. You could find short excerpts from a book online but not the

full text.

- 2005: The Authors Guild and the Association of American Publishers sued Google for violation of their copyrights.
- October 28, 2008: After arduous negotiations, Google and the plaintiffs filed a proposed settlement with the Southern Federal District Court of New York.
- November 13, 2009: In response to hundreds of objections filed with the court, Google and the plaintiffs submitted an Amended Settlement Agreement (ASA).
- February 18, 2010: Judge Chin conducted a fairness hearing at which more objections were raised.
- March 23, 2011: Judge Chin rejected the ASA.

What began as a project for online searching metamorphosed during those seven years into an attempt to create the largest library and book business ever imagined. Had Google kept to its original plan, it might have won its case by invoking the doctrine of fair use. To display a few sentences in the form of snippets could hardly be equated with reproducing so much text that Google was effectively appropriating the bulk of a book. The early version of Google Book Search did not amount to commercial competition with publishers, because Google provided its search service free of charge, although it linked its displays to advertisements.

Then the lawyers took over. For more than two years, the legal teams of Google and the plaintiffs wrangled over details of how their differences could be resolved by a partnership in a common commercial enterprise. (The lawyers' fees for the various parties eventually came to \$45 million.) The result, Google Book Search, had many positive aspects. Above all, it promised to provide millions of readers with access to millions of books. It also gave authors an opportunity to have their out-of-print works revived and circulated widely, instead of lying unread on the shelves of research libraries. The authors would collect fees from the retail sales of the digital copies, and the libraries would gain access to the entire data bank, consisting of millions of books, by paying an annual subscription fee. If the prices were moderate, everyone would benefit.

The settlement had many other advantages: free service on at least one terminal at public libraries, special measures to help the visually impaired, and access to Google's database for large-scale quantitative research. Its main disadvantage, according to many critics, was its commercial aspect. Google asked libraries to supply it with their books free of charge—not quite free, actually: Google paid for the digitizing but the libraries

shouldered heavy transactional costs. (Harvard paid \$1.9 million to process the 850,000 public domain books that it furnished to Google.) In return, the libraries were required to buy back access to those books in digital form for a subscription price that might escalate to a ruinous level. The subscription rate would be set by a Book Rights Registry composed of representatives of the authors and publishers who had an interest in maximizing their income. Therefore, the settlement could look like a way to conquer and divide a lucrative market: 37 percent of the income would go to Google, 63 percent to the plaintiffs, the authors and publishers who had become its partners. No one represented the public interest, and no public authority was empowered to monitor an operation that seemed likely to determine the fate of books far into the digital future.

In his opinion, Judge Chin did not dwell on the commercial aspects of Google Book Search, except insofar as they posed a threat to restrain competition. Two memoranda from the Department of Justice had alerted him to the danger of a violation of the Sherman Antitrust Act, and he especially objected to the way that threat applied to the digitization and marketing of “orphan” books—books whose copyright owners have not been identified. Orphan books—and unclaimed copyrights in general—are crucial to the entire enterprise, because there are so many of them, perhaps five million, according to a recent estimate. Most of them date from the period between 1923 and 1964, when copyright law is particularly ambiguous. Any database that excluded them would be disastrously deficient, but any enterprise that included them would expose the digitizer to ruinously expensive lawsuits. Damages would probably run to at least \$100,000 per title. The settlement solved this problem by giving Google exclusive exemption from litigation. If any owners of unclaimed copyrights identified themselves, they would be compensated, but they could not collect damages.

In its original version, the settlement went further. It made Google and the plaintiffs effective proprietors of the orphan books and permitted them to pocket the income from their sale, even though hardly anyone involved in Google's enterprise had ever had anything to do with the creation of those works. The amended version of the settlement eliminated that provision, but it continued to give Google exclusive legal protection in a manner that would discriminate against potential competitors. It amounted to changing copyright law by litigation instead of legislation.

In objecting to this aspect of the settlement, Judge Chin insisted that issues of such importance should be decided by Congress, all the more so since the settlement would determine future activities instead of merely remedying damages that took place in the past. Class action suits that affect the future look dubious in court, and the Google Book Search case also included a doubtful opt-out provision. It provided that any author of a book that was covered by copyright but no longer commercially available (that is,

essentially, in print) would be deemed to have accepted the terms of the settlement unless he or she explicitly notified Google to the contrary. Judge Chin noted that 6,800 authors had opted out, an indication that the settlement may not have looked acceptable to a considerable proportion of the class that the Authors Guild claimed to represent.

How large is that class? The Guild has 8,000 members, but there must be far more than 100,000 living writers who have published a book during the last fifty years. Many of them are academic authors who do not depend on the sale of books to make a living. Some of them sent memoranda to the court saying that they preferred to have their out-of-print books made available free of charge, because they cared more about the diffusion of their ideas than what little income they might derive from sales. Of course, professional writers have a vital interest in sales, and they understandably pressed hard to make the most from the deal with Google. Judge Chin did not disparage anyone's motives, but he showed concern for the representativeness of the class composed of authors that was involved in the class action suit and the antagonistic interests of different groups of its members.

Judge Chin also mentioned other problems that had been stressed in the five hundred amicus briefs and memoranda that had been submitted to the court. Two stand out.

Foreign authors and publishers objected that the settlement violated international copyright law. Google digitized many of their works without their permission, even though they held copyrights in their home countries. The settlement treated them as if they belonged to the same class as the American rightsholders, despite the fact that they had little possibility of studying the terms of the settlement and opting out of it. The ASA met most of those objections by eliminating copyrighted books that were published abroad, except in the United Kingdom, Canada, and Australia. But foreigners continued to protest about the potential violation of their rights and noted that they, too, had an orphan book problem.

To many who sent their objections to the court, as well as others, Google Book Search threatened to violate their privacy. In the course of administering its sales, both of individual books and of access to its database by means of institutional subscriptions, it would accumulate information about the private activity of reading. It would know who read what, including in many cases the precise passages that were read and the exact time when the readers consulted them. The ASA provided some assurances about this danger, but Judge Chin recommended more, should the ASA be revised and resubmitted to the court.

He also urged the possibility that a further revision of the settlement might be acceptable

to the court if its key provisions were switched from opt-out to opt-in requirements. In that case, presumably, the authors of copyrighted, out-of-print books would not be considered to have accepted the settlement unless they gave notice of their intention to do so. If enough of those authors could be located, or volunteered to consent to the settlement, Google Book Search might build up a large database of books published since 1923. But the logistics and the transaction costs might make that task unfeasible, and the problem of orphan books would remain unsolvable without congressional legislation.

The cumulative effect of these various objections, many of them endorsed by Judge Chin's decision, could give the impression that the settlement, even in its amended version, is so flawed that it deserves to be pronounced dead and buried. But that would mean the loss of its many positive features. How could its advantages be preserved without the accompanying drawbacks? The answer that I and others have proposed is to create a Digital Public Library of America (DPLA)—that is, a collection of works in all formats that would make our cultural heritage available online and free of charge to everyone everywhere.

Having argued so often for this alternative to Google Book Search, I may fall victim to the syndrome known in France as preaching for one's own saint. Instead of repeating the arguments previously made in these pages and elsewhere,^{*} I would like to show how the case for the Digital Public Library would look if seen from the perspective of similar projects in other countries.

The most impressive attempts to create national digital libraries are taking shape in Norway and the Netherlands. They have state support, and they involve plans to digitize books covered by copyright, even those that are currently in print, by means of collective agreements—not legalistic devices like the class action suit employed by Google and its partners, but voluntary arrangements that reconcile the interests of the authors and publishers who own the rights with those of readers who want access to everything in their national languages. Of course, the number of books in Norwegian and Dutch is small compared with those in English. To form an idea of what could be done in the United States, it is better to study another venture, the pan-European digital library known as Europeana.

Europeana—which already has offices in The Hague—is still in a formative phase, but its basic structure is well developed. Instead of accumulating collections of its own, it will function as an aggregator of aggregators. Information will be accumulated and coordinated at three levels: particular libraries will digitize their collections;



national or regional centers will integrate them into central databases; and Europeana will transform those databases, from twenty-seven constituent countries, into a single, seamless network. To the users, all these currents of information will remain invisible. They will simply search for an item—a book, an image, a recording, or a video—and the system will direct them to a digitized version of it, wherever it may be, making it available for downloading on a personal computer or a handheld device.

To deliver such service, the system will require not only an effective technological architecture but also a way of coordinating the information required to locate the digitized items—“metadata,” as librarians call it. The staff of Europeana at The Hague has perfected a code to harmonize the metadata that will flow into it from every corner of Europe. Unlike Google, it will not store digital files in a single database or server farm. It will operate as a nerve center for what is known as a “distributed network,” leaving libraries, archives, and museums to digitize and preserve their own collections in the capillary system of the organic whole.

A digital library for America might well follow this model, although Europeana has not yet proven that it is workable. When a prototype went live on November 20, 2008, it was flooded with so many attempts at searches that the system crashed. But that failure can be taken as testimony to the demand for such a mega-library. Since then, Europeana has enlarged its capacity. It will resume functioning at full tilt in the near future; and by 2015 it expects to make thirty million items, a third of them books, available free of charge.

Who will pay for it? The European Union will do so, drawing on contributions from its member states. (Europeana’s current budget is e4,923,000, but most of the expenses fall on the institutions that create and preserve the digital files.) This financial model may not be suitable for the United States, but we Americans benefit from something that Europe lacks: a rich array of independent foundations dedicated to the public welfare. By combining forces, a few dozen foundations could provide enough money to get the DPLA up and running. It is impossible at this point to provide even ballpark estimates of the overall cost, but it should come to less than the e750 million that President Sarkozy pledged for the digitization of France’s “cultural patrimony.”

Moreover, in building up its basic collections, it could draw on the public-domain books that are currently stored in the digital archives of not-for-profit organizations like Hathi Trust and the Internet Archive—or (why not?) in the servers of Google itself, Google willing.

Once its basic structure has been erected, the Digital Public Library of America could be enlarged incrementally. And after it has proven its capacity to provide services—for

education at all levels, for the information needs of businesses, for research in every conceivable field—it might attract public funds. Long-term sustainability would remain a problem to be solved.

Other problems must be confronted in the near future. As the Google case demonstrated, nearly everything published since 1923, when copyright restrictions begin to apply, is now out of bounds for digitization and distribution. The DPLA must respect copyright. In order to succeed where Google failed, it will have to include several million orphan books; and it will not be able to do that unless Congress clears the way by appropriate legislation. Congress nearly passed bills concerning orphan books in 2006 and 2008. It failed in part because of the uncertainty surrounding Google Book Search. A not-for-profit digital library truly devoted to the public welfare could be of such benefit to their constituents that members of Congress might pass a new bill carefully designed to protect the DPLA from litigation should rightsholders of orphan books be located and bring suit for damages.

Even better, Congress could create a mechanism to compensate authors for the downloading of books that are out of print but covered by copyright. In addition, voluntary collective agreements among authors of in-print books, similar to those in Norway and the Netherlands, could make much contemporary literature accessible through the DPLA. The copyright problems connected with works produced outside the United States might be resolved by agreements between the DPLA and Europeana as well as by similar alliances with aggregators on other continents. Items that are born in diverse formats such as e-books pose still more problems. But the noncommercial character of the DPLA and its commitment to the public good would make all such difficulties look less formidable than they seemed to be when they were confronted by a company intent on maximizing profit at the expense of the public and of its competitors.

In short, the collapse of the settlement has a great deal to teach us. It should help us emulate the positive aspects of Google Book Search and avoid the drawbacks that made Google's enterprise flawed from the beginning. The best way to do so and to provide the American people with what they need in order to thrive in the new information age is to create a Digital Public Library of America.

1. *

See, for example, "[Can We Create a National Digital Library?](#)," *The New York Review*, October 28, 2010. ↩

© 1963-2011 NYREV, Inc. All rights reserved.