newyorker.com

Google Books and the Judge

by ANTHONY GRAFTON • SEPT. 18, 2009

For years, Google has kept rolling out one new application after another, from Street View to Fast Flip, and rolling over most of the competition as it does so. They've just done it again: yesterday, Google announced that they would allow On Demand Books to produce paperback versions of two million out-of-copyright digitized texts on the Espresso Book Machine, which can print and bind a book in less than five minutes. For the recommended $8 a copy—the price of a paperback romance—readers will be able to buy anything from "Moby Dick" to "Dame Curtsey's Book of Candy Making," The Google-On Demand partnership could transform retail bookselling—especially of books for university courses.

Google's decision to feed its library into the Espresso Book Machine (and to donate its own share of revenue to charity) is not simply a sign of its power; the Google Books project is facing a crunch point. In 2005, groups of publishers and rights holders filed suit, claiming that digitizing works and allowing partial access to them without permission violated copyright. Long negotiations have led to a proposed settlement, to be ruled on by Judge Denny Chin. Concerned parties are making their submissions to the court; the deadline is today.

The settlement has a lot to offer most ordinary authors—those of us whose books sell in the high hundreds or low thousands, and then go out of print. Google will pay sixty dollars for every book for which it can find a rights holder and will share any future revenues with authors and publishers. More important, millions of books that are in copyright but out of print (and hard to find) will get another chance. People searching for information will learn from Google that these books exist and then be able to read sections of them online. The system will provide immediate links to libraries where the full texts can be found and to retailers, if any, who sell them. Any rightsholder who doesn't want to take part can opt out. From most writers' standpoint it looks like a decent deal.

The closer you look, though—and a lot of people, from government and Google's competitors to its former collaborators, are looking closely—the more problems you see. The Justice Department will probably submit an opinion, and it's likely to be critical, based on what Marybeth Peters, the U.S. Register of Copyrights, told the House Judiciary Committee. France and Germany have complained that Google is violating their copyright laws. Google, which has digitized thousands of books from both countries, may have to hide them from European users—and Europeans see offering these books to Americans as an infringement of their rights. Amazon has already denounced the settlement, predictably but eloquently, as licensing an uncontrollable monopoly. So have the trustees of many writers' estates, who reasonably fear that the settlement would harm the copyright protections that ensure income streams.

Even the libraries that have provided Google with its raw materials are not all happy with the result. The out-of-print books Google has digitized come from nonprofit institutions that built their collections as a public good. In return for pocket change—Google will contribute $125 million to create a nonprofit rights registry—these public treasures will now be monetized for the benefit of a private corporation. True, Google will give every public and university library one terminal where readers can access its entire collection. But these machines won't be able to download or print texts—and you can imagine the lines. Libraries that want full access to all the books in Google will have to pay for the privilege, as well as for every download.

Google, with its mission "to organize the world's information and make it universally accessible and useful," plans to turn itself into the biggest bookstore the world has ever known, and to make libraries pay for acting as its agents. It's troubling that the libraries that already have the richest collections will also be the ones that can offer their users the full Google service. Harvard was one of Google's original partners. But Robert Darnton, Harvard's librarian and a fan and creator of digital projects, has ceased supplying books still in copyright to Google. As he has written, "To digitize collections and sell the product in ways that fail to guarantee wide access ... would turn the Internet into an instrument for privatizing knowledge that belongs in the public sphere." Other Google partner libraries, however, support the settlement and have criticized Darnton's decision.

No judge can decide the biggest question here, the one I explored in *The New Yorker* back in 2007: what kind of library is Google actually building? It's certainly enormous. Google states that users can now search "the full text of some seven million books." Some report that the system as a whole already contains ten million. Either way, Google is on its way to surpassing the 13.8 million volumes of the British Library, which has been collecting in one form or another for centuries, and, eventually, even the thirty-two million in our Library of Congress. And unlike those vast, material collections, imprisoned in great buildings, Google's is everywhere.

But Google has misdated hundreds of books and scattered many multivolume works so arbitrarily that they're hard to piece together even with the computer's help. Suppose you want that first edition of Middlemarch, which came out in eight parts in 1871-72. If you enter "Eliot Middlemarch 1871," you'll find parts 1 and 2; if you enter "Eliot Middlemarch 1872," you'll get parts 5 and 6; and you can probably enter something else, though I haven't yet found out what, to reach parts 3-4 and 7-8. But the system never tells you that this is the first edition, or how you can put it all together. And bizarrely, Google sorts books, as Geoffrey Nunberg and others have shown, not by the Library of Congress Classification, but by the Book Industry Standards and Communications used by publishers to tell booksellers where to stow a given item—usually a celebrity memoir or a cook book. That explains why, as Geoffrey Nunberg pointed out in the *Chronicle of Higher Education*, "the 2003 edition of Susan Bordo's Unbearable Weight: Feminism, Western Culture, and the Body (misdated 1899) is assigned to Health & Fitness."

Google excels at listening to users, identifying complaints, and fixing them. They'll correct mistakes users bring to their attention and will continually improve the system as a whole. But it's utopian to believe that the company could or would repair the millions of errors already built into the system—or that new problems won't continue to crop up, as Google vacuums up more millions of books without finding out in advance what book professionals know about how best to identify and organize them.

Will the juggernaut keep rolling? We'll know later this year. But should it? It may be too late if and when we find out.

Original URL:

http://www.newyorker.com/online/blogs/books/2009/09/google-books-and-the-judge.html