A universal digital library is within reach

Copyright law poses considerable challenges, but any barriers to mass digitization of the world’s books can — and should be — overcome.

By Pamela Samuelson

May 1, 2012

Since 2002, at first in secret and later with great fanfare, Google has been working to create a digital collection of all the world’s books, a library that it hopes will last forever and make knowledge far more universally accessible.

But from the beginning, there has been an obstacle even more daunting than the project’s many technical challenges: copyright law.

Ideally, a digital library would provide access not only to books free from copyright constraints (those published before 1923), but also to the tens of millions of books that are still in copyright but no longer in print.

Copyright law makes it risky to digitize these books without permission from copyright owners, and clearing the rights can be prohibitively expensive (costing on average, according to estimates, about $1,000 per book). Even if the money wasn’t a problem, hundreds of thousands — and probably millions — of books are likely to be "orphan works" whose rights-holders are unknown or can’t be found.

Google bumped up against copyright law in 2005, when lawsuits were filed by the Authors Guild and by a group of five publishers alleging that Google’s scanning of books from major research library collections constituted copyright infringement. Google argued that scanning books to index their contents and make snippets available online was fair use, not infringement. But with its potential liability running into the billions or even trillions of dollars, Google was understandably receptive to overtures from the Authors Guild and publishers to settle instead of litigate.

A settlement announced in October 2008 would have given Google a license to keep scanning books and to help re-commercialize those that were out of print by running ads
next to search results, selling books to consumers, and licensing digitized out-of-print books to libraries and other institutions. Google could have also displayed up to 20% of a book's contents in search queries.

But the proposed settlement fell apart in March 2011 when the judge overseeing the case ruled that it was unfair to the authors and publishers on whose behalf it had been negotiated, and that it would give Google "a de facto monopoly over unclaimed works." The proposal went far beyond the issues in litigation, he concluded, addressing matters "more appropriately decided by Congress" than through litigation.

But the dream of a universal digital library lives on. Now a coalition of libraries and archives has come together to create a Digital Public Library of America to fulfill the original vision of a digital library for all. It could well be that an effort without commerce in the mix will have an easier time of it.

A broad consensus already exists to remove copyright obstacles to orphan works. There is also growing interest in mass digitization of out-of-print works. The arguments for increased access are compelling: These books aren't producing any revenue for copyright owners, and most of them are unlikely to be reprinted. Libraries already own copies of many of them and want to make them available digitally to their communities. And rights holders can always opt out of a library mass-digitization project.

The U.S. Copyright Office recognizes that barriers to mass digitization need to be overcome. It proposed a partial legislative fix, which became the Orphan Works Act of 2008. The bill passed in the Senate, but then stalled in the House. Maria Pallante, who heads the Copyright Office, recently announced her interest in renewing this legislative initiative at a Berkeley Law conference on orphan works.

Meanwhile, the European Union is also working on a legal framework to allow greater access to orphan works. France has adopted legislation permitting libraries to mass-digitize books that aren't in print but are still in copyright. Germany is considering a similar proposal. Japan and Norway have authorized national libraries to undertake mass-digitization projects that even include in-copyright works. The U.S. should not lag behind.

Digital libraries containing millions of out-of-print and public domain works would vastly expand the scope of research and education worldwide, extending access to millions of people in undeveloped countries who don't have it now. It would also open up amazing opportunities for discovery of new knowledge. Being able to conduct searches over a corpus of millions of books allows researchers to learn things never before possible.

There are three promising strategies for removing barriers to a universal digital library: First, it should be considered "fair use" in copyright law for nonprofit libraries to circulate orphan works for their patrons for noncommercial purposes. Second, Congress should pass legislation to limit damages and injunctions for other reuses of orphan works. Third, the Copyright Office should explore a collective licensing program under which all
in-copyright but out-of-print works could be made available, as some countries are now trying.

Workable solutions exist to fulfill the dream of a universal digital library. Do we really want to tell our grandchildren that we could have achieved this goal but lacked the will to do so?

Pamela Samuelson is a professor at the UC Berkeley School of Law and faculty director of the law school's Berkeley Center for Law & Technology.

Copyright © 2012, Los Angeles Times