July 9, 2012

Reforming Copyright Is Possible
And it's the only way to create a national digital library

By Pamela Samuelson

The tantalizing vision of universal access to the cultural and scientific heritage of humanity seemed close to fulfillment in 2008, when Google announced the settlement of a class-action lawsuit charging that its Google Book Search project infringed copyright by scanning in-copyright books from major research-library collections. Sergey Brin, co-founder of Google, later asserted that the settlement was a big win for everyone because it would create a "library to last forever."

But it was not to be. The very ambitiousness of the settlement was its undoing. In 2011 a federal judge ruled against it, mainly because it went too far beyond the issues in litigation, which concerned only whether scanning books to index their contents and make snippets available was infringement or the limited exception, fair use, since snippets would not supplant—and might enhance—demand for the works. Having failed to reach a more limited settlement, the litigants are expected to go to trial this fall.

The failure of the Google Book settlement, however, has not killed the dream of a comprehensive digital library accessible to the public. Indeed, it has inspired an alternative that would avoid the risks of monopoly control. A coalition of nonprofit libraries, archives, and universities has formed to create a Digital Public Library of America, which is scheduled to launch its services in April 2013. The San Francisco Public Library recently sponsored a second major planning session for the DPLA, which drew 400 participants. Major foundations, as well as private donors, are providing financial support. The DPLA aims to be a portal through which the public can access vast stores of knowledge online. Free, forever.

Initially the DPLA will focus only on making digitized copies of millions of public-domain works available online. These include works published in the United States before 1923, those published between 1923 and 1963 whose copyrights were not renewed, as well as those published before 1989 without proper copyright notices,
and virtually all U.S.-government works.

If a way can be found to overcome copyright obstacles, many millions of additional works could be made available.

It's no secret that copyright law needs a significant overhaul to adapt to today's complex information ecosystem. Unfortunately the near-term prospects for comprehensive reform are dim. However, participants at a conference last spring at Berkeley Law School on "Orphan Works and Mass Digitization: Obstacles and Opportunities" believe that modest but still meaningful reforms are possible.

As keynote speaker, Maria Pallante, register of copyrights, announced her intent to move forward with a legislative initiative to foster uses of so-called orphan works (those whose rights holders cannot be located after a diligent search), to update rules that allow libraries to make reasonable uses of in-copyright works, and to explore ways to enable mass digitization of works that are unavailable commercially. She and others noted that several countries have recently authorized large-scale mass-digitization projects. The French Parliament, for instance, has enacted a law to allow the national library to digitize out-of-commerce works—those no longer commercially available—in its collection.

One of the key problems for digital libraries such as the DPLA is the extraordinarily long terms of copyright today: 70 years past an author's death or 95 years from first publication for works made for hire. Had Congress not bowed to pressure from industries with a stake in copyright, especially Hollywood, to extend existing copyright terms several times in the past 40 years, all works published before 1956 would now be in the public domain and available for inclusion in the DPLA.

Because of term extensions, copyright is a significant impediment to reuse of most books published in the 20th century, even though an overwhelming majority have long been out of print. That puts DPLA planners in a tough position. They want to provide public access to literatures of the century just past, but it is impractical to clear rights on a work-by-work basis. (It would cost an average of about $1,000 per work, and that's not counting any royalty payments.)

For millions of orphan works, rights cannot be cleared at all, because the owners cannot be found. The problem is especially troublesome for special collections of historically significant materials, for which copyright information is often unavailable. At the Berkeley conference, Lydia Loren, a professor at Lewis & Clark Law School, argued that such works are hostages in need of
someone who cares enough to free them from copyright bondage, not orphans that should be consigned to a bereft future.

The U.S. Copyright Office recognized the severity of the problem in 2006 and recommended legislation to allow free reuses of copyrighted works whose rights-holders could not be found after a reasonably diligent search. The Senate passed a bill similar to the copyright office’s proposal in 2008. However, owing in part to the announcement of the Google settlement, which proposed to give Google a license to make orphan works available to the public, further action on the bill stalled. Pallante’s recent announcement that the copyright office will renew efforts to deal with orphan works is therefore good news.

Still, legislation may not be needed for some reuses of orphan works. Jennifer M. Urban, director of the Samuelson Law, Technology & Public Policy Clinic at Berkeley, has suggested that fair use can be part of the solution. Many orphans that nonprofit libraries want to make available online are fact-intensive works, written by scholars for scholars, whose motivation is to share knowledge. Insofar as such works are unavailable commercially and their rights-holders cannot be located, there is no risk of harm to any existing or potential market for the works, an important factor in fair-use cases.

Although copying the whole of protected works generally cuts against fair use, the modern trend in copyright cases is to ask whether the amount taken was reasonable in light of the putative fair user’s purpose. Nonprofit libraries can argue that they are providing access to orphan works in order to enable research, scholarship, and teaching, all three of which are statutorily favored uses. That conforms to the recent ruling in Cambridge University Press et al. v. Becker et al. that nonprofit educational purposes strongly favor fair use. In that case, Georgia State University prevailed in defense of its electronic course-reserve policy against a publisher challenge.

Fair use of orphan works may also be appropriate if a library keeps good and accessible records about its diligent search for the owner/author and establishes a waiting period between announcing its belief that a work is an orphan and making the work available.

Fair use has long been a flexible balancing tool enabling copyright law to adapt to developments unforeseen by Congress. No one in 1976, for instance, imagined search-engine technologies or the World Wide Web. Yet fair use enabled several search engines to fend off copyright lawsuits in justifying copying the contents of
open Internet sites, indexing Web contents, and serving up links to or thumbnail images of those contents in response to queries. Such precedents very likely emboldened Google to rely on fair use to justify mass digitization for indexing book contents and making snippets available to its search-engine users. If Google wins its fair-use defense this year, the precedent would help library digitization projects.

University libraries are rooting for Google's fair-use defense because it would also bolster their fair-use justification for possessing and using library digital copies, or LDC's, provided by Google of books from their collections. The digital copies are an important way to preserve library resources so they will be available for lawful uses today and future generations tomorrow.

The LDC's are also valuable to enable scholars to engage in research that was previously unimaginable, like tracing the influence of a philosopher over time by finding all references to him in the LDC corpus. Because text mining makes only computational uses of in-copyright works, there is a strong argument that it is fair use.

One of Google's library partners, the University of Michigan, claimed fair use when it and HathiTrust, repository of Michigan's and some other Google partners' LDC's, initiated a program to conduct searches for possible rights-holders of orphan works. The Authors Guild has sued Michigan, several other university libraries, and HathiTrust, alleging that libraries have no fair-use privileges beyond the contours of the special library-use exception in U.S. law, which the guild asserts does not permit mass digitization. Fair use may be put to a very important test in this case as well as in the Google Book Search case.

Yet the suits against Google and Michigan may well falter, because it is unclear whether the plaintiffs have standing. After all, the Authors Guild owns no copyrights that the LDC's allegedly infringe, and it represents only a very small proportion of published authors in the United States. (The OCLC Online Computer Library Center, operated by a nonprofit library association, estimates that some 3,685,778 personal and 977,679 corporate authors have published books in the United States since 1922; in contrast, the Authors Guild has some 8,000 members.) Google also fought the class action by asserting that the individual guild-member plaintiffs did not adequately represent the interests of the vast majority of authors whose books Google has scanned. A survey it commissioned shows that only a small minority of published authors objected to scanning snippets or believed that doing so would harm them. However, Judge Denny Chin was not persuaded by the standing
challenges and has certified the class for trial.

Even if Google and Michigan win their fair-use defenses (and they may not), that would still not free up millions of non-orphans for inclusion in a universal digital library. The victories would, however, strengthen the argument that libraries should be able to make available as fair use long-out-of-commerce works (for example, books published in the 1960s that have been commercially unavailable for 40 years). Additional progress is possible through efforts to encourage copyright owners to make their out-of-print works available to the DPLA through licenses from the nonprofit Creative Commons or the like. Though such progress would be welcome, a universal digital library would be slow to come into being through those steps.

The fastest way to achieve a more comprehensive digital library is for Congress to create a license so that digital libraries could provide public access to copyrighted works no longer commercially available. This approach would make it unnecessary to engage in costly work-by-work searches for rights holders and would free up orphan works. A digital library such as the DPLA could pay a fee for a license to display such works to the public for noncommercial uses. Rights-holders could come forward to get compensation for the uses of their works, or they could opt out. Works whose rights-holders failed to show up within a certain period of time (perhaps five or 10 years) could be presumed to be orphans and made available on an open-access basis.

At an abstract level, the licensing option seems attractive. It resembles the basic framework of the recently enacted French law authorizing the national library to digitize out-of-commerce works. However, France has at least three advantages over the United States.

For one thing, the national library is an existing entity. It has a half-million volumes in its collection that can be digitized under the law. Although the Library of Congress has more than 32 million books in its collection, it has not yet exhibited interest in mass-digitizing them. The DPLA might be a potential licensee, but it is a work in progress without a collection of its own.

Or perhaps Google could be persuaded to offer a copy of the Book Search corpus to a digital-library project as part of a legislative effort to shield it from the billions of dollars in statutory damages that might be imposed if the Authors Guild wins its lawsuit.

Another alternative is for Congress to authorize a public-private partnership to create a digital public library whose corpus would
also be available to other search engines besides Google to improve their search capabilities. A 10-million-book digital library might cost $100-million, but that is a modest amount if it gives the public access to a vast repository of cultural and scientific resources.

A second advantage for France is that it has existing collective-management organizations to serve as appropriate licensors of copyrighted resources. Those societies already have thousands of rights-holders as members to whom they already pay royalties. The closest approximation in the United States is the American Society of Composers, Authors and Publishers, but it licenses only public performances of music.

The Copyright Clearance Center might initially seem a possible institutional licensor, as it has relationships with many publishers for which it collects fees for licensing photocopies of textual works. But it has a far more limited role in licensing than the typical collective-management society, and it represents only a fraction of the rights-holders whose works would be licensed under mass-digitization legislation.

Libraries would be unlikely to support a collective-licensing approach without firm commitments about the reasonableness of prices and terms over time and assurances that the system would not undermine their right to rely on fair use when appropriate. A framework for good governance of a collective-licensing regime and for government oversight for possible abuses would probably also be necessary to build confidence in this solution.

At the moment in the United States, there is unfortunately no obvious candidate available to designate as the licensor of a Congressionally authorized mass-digitized corpus.

A third advantage for France is that major publishers' and authors' groups supported the legislation to digitize works in the national library. Although the Authors Guild and the Association of American Publishers reached agreement to license mass digitization and public displays of books under the Google Book settlement, there was considerable dissension among other groups of authors and publishers.

While libraries supported the settlement, they were vocal about their concerns that institutional subscriptions to the Google corpus of out-of-commerce works would rise to excessive levels over time. It may, therefore, be difficult to get consensus in the United States that would enable similar legislation to pass.

After the Google Book settlement failed, I outlined a legislative
package in the *Columbia Journal of Law & the Arts* that would bring about virtually all of the benefits envisioned by proponents of the settlement without the downsides of a Google monopoly. In brief, I recommended: 1) creating a privilege to scan in-copyright works for preservation, indexing, and text mining of the works; 2) allowing orphan works to be made available on an open-access basis; 3) expanding the right of libraries and others to improve access for those who have trouble reading print; 4) ensuring that reader-privacy interests are respected.

I also suggested that serious consideration be given to creating an extended collective-licensing regime for out-of-commerce, non-orphan books so that an institutional subscription database of those works might be created. Such licenses have been used with considerable success in Nordic countries. Extended licenses provide rights-holders with compensation while at the same time allowing users to get a license to make a large number of works available when the transaction costs of clearing all rights, one by one, would be excessive or impossible. It is worrisome, however, that experience with collective licensing in Australia has shown that even if prices and terms are reasonable at the outset, they may incrementally rise to unreasonable levels as time goes on. Checks and balances would need to be carefully built in to any U.S. collective-licensing regime.

Copyright law needs considerably greater reform than the measures I’ve discussed. Copyright should be shorter in duration, more balanced, more comprehensible, and normatively closer to what members of the public think that it means or should mean.

Although we are not likely to get comprehensive reform anytime soon, perhaps we can persuade Congress to make some more modest reforms.

We know it is now possible for the cultural and scientific heritage of humankind to be made available through a universal digital library such as the DPLA. It would be a grievous mistake not to bring that future into being when it is so clearly within our grasp.

*Pamela Samuelson is a professor of law and of information at the University of California at Berkeley and faculty director of the Berkeley Center for Law & Technology.*