



Great New Possibilities for the Library of Congress!

Citation

Darnton, Robert. 2015. "Great New Possibilities for the Library of Congress!" New York Review of Books, August 13, 2015.

Published Version

<http://www.nybooks.com/articles/2015/08/13/great-new-possibilities-library-congress/>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:34307198>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

The prospect of new leadership at the Library of Congress opens up great possibilities for democratizing access to culture. America has led the world in making digital communication part of everyday life, yet it has lagged behind other countries in making the holdings of its greatest library available online and free of charge. A new regime at the Library of Congress could digitize its collections and link them with collections in other libraries, archives, and museums so that everyone has access to the resources that are everyone's heritage.

Until now, digital development has taken place primarily in the private sector. While corporations such as Google were dominating the Internet, the public interest in the digital realm was left to private initiative. The Digital Public Library of America, based in Boston, has begun to link the collections of research libraries in a national network, which now makes 10 million items available free of charge to everyone with access to the Internet. The Internet Archive, with headquarters in San Francisco, performs a similar service by harvesting texts from millions of websites as well as books. HathiTrust, located in Ann Arbor, Michigan, preserves the texts of 13.4 million volumes, largely from collections that were digitized by Google. (Google's own data base cannot be made available as a commercial digital library, owing to a decision of a federal district court, which declared Google Book Search illegal in 2010.)

Each of these initiatives provides valuable services. Although they overlap in places, they should be maintained—and above all, they should be integrated in a single system so that everyone has access to all of the country's cultural resources. The Library of Congress is the richest resource of all. As it is supported by public funds, the public should be able to tap its collections—106 million items, of which 37 million are books and other print materials—by means of the Internet.

Easier said than done. But other countries have done it, at least on a small scale. National libraries throughout Scandinavia have developed systems to make all of their collections available to all of their citizens, under various restrictions. They favor “extended collective licensing” agreements, backed by legislation, which make it possible for Scandinavians to access copyrighted material without the assent of every owner of a copyright. Associations of authors and publishers agree to the establishment of a fund and a collecting agency, which provides payments to the rights holders according to a formula such as a fixed fee for every

viewing of a page. Thanks to an arrangement of this sort worked out between the Norwegian National Library and a collecting society called Kopinor, Norwegians can read online nearly all the books that were ever published in their language.

The National Library of Australia operates a successful online system called Trove, which connects users with libraries scattered across the country. Australians click onto Trove and then follow the links that it provides, gaining access to 18.5 million books and theses, 155 million articles and reports, and 7.8 million photographs and other images. They cannot download the material covered by copyright, but they can borrow it by interlibrary loan. The National Library therefore functions as an aggregator—that is, a provider of catalogue-type information (metadata), which integrates all of Australia's resources into a single system. It is also digitizing its own collections, including most of Australia's newspapers.

The National Diet Library of Japan, which is similar to the Library of Congress, began digitizing its collections in 2009 after a revision of Japan's copyright law. Most of them are now available from its website, although books protected by copyright can only be consulted from computers located within the library itself.

South Korea has adopted a similar policy, designed to create an "IT Nation." Its national library began digitizing material in 1998, and eleven years later it completed the construction of a "National Digital Library," which makes its holdings accessible online, except for copyrighted works, which must be read on site.

Having made a large proportion of its historical collections available through a digital repository called Gallica, the Bibliothèque nationale de France (BnF) is now digitizing works from the twentieth century, including many that are covered by copyright. True, it did not see much of the \$1.1 billion that President Sarkozy pledged in 2009 to make the country's cultural heritage accessible online. But legislation passed on March 1, 2012 empowered it to create a freely accessible data base from copyrighted works that were published before January 1, 2001 and that are no longer being commercially distributed. The rights owners will be compensated for ten years by a collective management organization representing

authors and publishers. They can opt out of the arrangement, and they may sign up with Google for the sale of their out-of-print works; so it is not yet clear how the French will have access to twentieth-century literature. But the BnF is committed to accessibility, and it has also developed a solution to the problem of orphan works—books under copyrights of which the owners have not been identified. The BnF will post the titles of such books from its collections; and if no claimant comes forward within ten years, it will digitize the texts and make them available free of charge.

Publishers in Britain are required by law to provide a free copy of every book they publish to the British Library and five other “deposit libraries.” Since April, 2013, that requirement covers a great deal of digital material, including websites, blogs, and electronic journals. Legal **constraints** mean that access to copyrighted texts is restricted to computers within the library. Only one person at a time can consult the text of an e-book, and that person cannot print out more than one of its chapters or more than 5 percent of an article in a journal. Still, the British Library is committed to defending the public interest in the digital sphere: “Our mission,” it says, “is to make our intellectual heritage accessible to everyone.”

The Library of Congress has the same mission. But while other great libraries were leading the way into the digital future, it failed to manage its own information technology, to say nothing of developing a national network of electronic resources. How can the new librarian, who will assume office on January 1, achieve this goal?

Although it provides services to members of Congress, the Library of Congress is above all a national library. It is the only deposit library in the country, and it should acquire and preserve everything of importance that is published in the United States, “published” being taken in its original sense of “making public,” whatever the form, content, or medium may be.

Digital documents are more fragile than texts printed on paper, because the minute ones and zeros of which they are composed easily unravel; and even if they resist [**ed: “physical” is deleted here**] damage such as “link rot,” they can get lost in cyberspace. Without up-to-date identifying descriptions (“metadata”), they

might as well not exist. The danger of disappearing into electronic clouds is compounded by inconsistencies in the metadata furnished by every institution with a digital repository. Metadata must be adapted, or “scrubbed,” so that the material they identify can be linked in a seamless system of systems. The Library of Congress should set metadata standards and promote interoperability on a national scale; and while fostering advanced technology in the U.S., it should promote innovations that are compatible with the systems of other countries. It should collaborate with Europeana, an aggregator funded by the European Union, which links together data bases scattered throughout 28 countries in Europe.

Does this standardizing function mean that the Library of Congress would crush other libraries with its sheer size and weight? Not at all. Not any more than its earlier cataloguing service compromised the independence of other institutions. The Library of Congress should facilitate, not dominate the activities of libraries outside Washington. Instead of operating from the top down, the national network should function horizontally. The Digital Public Library of America **[ed: delete “(DPLA)”** has demonstrated the effectiveness of a horizontal or distributed system. It has no data base of its own, and it employs only thirteen staff members in its headquarters, but soon it will have “service hubs” in every state. Highly trained professionals in these hubs help public libraries develop their own digital resources. Librarians in small towns and urban neighborhoods invite local citizens to bring in photograph albums, letters, and family papers of all kinds. Under guidance from the hub’s experts, they scan this material, provide it with metadata, and preserve it. The collection grows organically, stimulating the community’s consciousness of its culture and history, and it is integrated in the national network of the DPLA. Public libraries are the most vital and most trusted institution in many communities. They could be empowered, not overpowered, by help of this kind from the Library of Congress.

What sets the Library of Congress (LOC) apart from every other library is the size of its collections, the largest in the world. They should be digitized. That, too, raises all kinds of difficulties, which, however, **could be eliminated by Congressional action and could at least** be minimized by careful study, including research on the extent of duplication among the collections of the LOC, HathiTrust and Google’s data base. Perhaps Google could be persuaded to donate copies of its files or to digitize everything in the LOC, while correcting the imperfections of its

earlier work. But any partnership between a government body devoted to the public welfare and a private corporation dedicated to maximizing profits is likely to run into trouble. (Google's arrangements to digitize collections in the Bavarian State Library and the Municipal Library of Lyon have aroused some fierce criticism.)

With adequate funding, the work could be done by the LOC itself or by the Internet Archive, a non-profit organization with extensive experience in mass digitizing. Brewster Kahle, the founding director of the Internet Archive, says that he digitizes at 10 cents a page or \$30 for an ordinary book and that he could get the cost down to 6 cents a page if he were to take on a very large library. At the higher rate, it would be possible to digitize all the books in the LOC for \$1.1 billion or about \$100 million a year for ten years. That is an enormous sum, but to put it in perspective one **should consider that South Korea paid \$108 million a year for seven years to build its National Digital Library. And one** should remember that the appropriation for the LOC in fiscal 2016 was \$631 million and that the Defense Department's budget request for fiscal 2015 came to \$575 billion (including \$79 billion for Overseas Contingency Operations).

It would be naïve to expect Congress to finance such a project. Funding for the LOC has been cut by 8 percent since 2010, but a coalition of foundations could provide \$100 million a year, and the benefit to the public would be incalculable—access to everything in the world's greatest library, instant information for all sorts of enterprises, and endless educational uses, especially if the material were to be accompanied by pedagogical packages and user-friendly data sets.

Whether the LOC digitized its own collections or coordinated access to those in other research libraries, it could not make available books, photographs, music, films, or anything else covered by copyright. Copyright now extends **automatically** to everything published after 1964, to many books published between 1923 and 1964, and even to some works that appeared as long ago as 1879. It therefore would be impossible for a fully digitized Library of Congress to make available most of the literature from the twentieth century—that is, most of its holdings. How could it cope with this problem?

The new administration of the LOC could learn from the experience of national libraries in other countries, especially in Scandinavia, where authors and publishers cooperate for their own benefit as well as that of the public. Most books, especially in the field of light fiction, rarely produce revenue after a year. Once their royalties have ceased, most authors have one remaining desire: to reach readers. I myself published a book in 1968 that brings in enough for me to take my wife out to dinner once every five years—if she pays her half of the check. With the agreement of my publisher, I have turned over the use of its copyright to an organization called Authors Alliance, which will make it available online and free of charge through a Creative Commons license. The vast majority of authors would gladly enter into an arrangement of this kind in order to give their books new life after they cease to produce any income.

Given its influence at the summit of the world of books, the Library of Congress could organize such a program. It could be entirely voluntary, or it could take the form of an extended collective licensing agreement with an opt-out provision for authors and publishers who prefer to remain attached to the market place. The new regime could create a moving wall so that every book, except those of the opt-outs, would fall into the public domain after ten years. Funding could be provided by modest pay-per-view fees, but it would be preferable to make the entire corpus available free of charge. An open-access system could include a mechanism for payments to authors and publishers according to the rate of accessing the texts—that is, reading them on a screen—without downloading and printing them.

This system [ed: “**probably**” **deleted**] would require [ed: “**Congressional**” **deleted**] legislation, **which would codify a tendency that already exists in** case law. Drawing on the principle of fair use, courts sometimes favor the diffusion of texts by non-commercial organizations for the benefit of the public, provided that private interests do not suffer any “market damage.” Music, photographs, and videos pose special problems; and books no longer really go out of print, since their texts are preserved in digital files, and they can be physically reproduced by the technology of print on demand.

The legal constraints on publishing are endlessly complex, and lawyers will find countless reasons to oppose innovation; but the pettifoggery should be

subordinate to the principles established by the Constitution, Article 1, Section 8, Clause 8: “To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.” The original copyright act of 1790, passed “for the encouragement of learning,” set the time limit at fourteen years, renewable once. The copyright act of 1998 extended the limit to the life of the author plus 70 years—effectively, more than a century. We need to revise current copyright law.

That may be impossible, considering the power of lobbies, especially those of Hollywood. But if the Library of Congress, which includes the Copyright Office, cannot lead the way to **alleviating** copyright restrictions, it can help solve the vexed problem of orphan works. It could establish a register of every book in its collection that, after due diligence, appears to be an orphan. It could publish the list of titles on its website. If anyone presents a valid claim to the rights of a book, it should remove that work from the list. And if no claimants appear within ten years, it should digitize all the books and make them available to the public from an open-access repository. The new librarian of Congress should create the repository, recruit the best experts to maintain it while developing other digital initiatives, and eventually fill it with the entire collection of the nation’s greatest library.

The repository of the LOC would then serve as the heart of a digital circulatory system which would energize the entire country. That prospect may raise so many difficulties that it looks utopian. But the founders of our country combined utopianism with a strong strain of pragmatism. By making the most of twenty-first century technology, the Library of Congress could tap the tradition of the founders and realign itself with its original principles.