Digitize, Democratize: Libraries and the Future of Books

by Robert Darnton

Abstract

Restrictive qualitications for entry, payment to obtain a reader's card, a general atmosphere of intimidation are some of the features of the existing great libraries; but they are also the betrayal of the very purpose of their establishement during the Modern Age, namely that of democratizing access to knowledge and circulating ideas. A goal which the Digital Public Library of America (DIPLA) project aims to achieve thanks to new technological infrastructures. It deals with an horizontal network growing day by day, that links up digital collections in libraries, archives and museums in the world designed with the aim to making the cultural heritage of america available, free of charge, to everyone in the world.

Keywords: Digital Public Library of America, diffusion of knowledge, cultural heritage, open-access, technological infrastructure.

Despite a lot of loose talk about the death of the book and the obsolescence of libraries, books and libraries are more important than ever in the current digital environment; and their importance will increase as we design the digital future – if only we can get it right. I want to develop this argument by looking backward into history and forward into plans to create a Digital Public Library of America.

The historical importance of libraries may seem obvious. At the University of Coimbra, the library embodies centuries of learning. And if you take a tour of a typical American university, you will notice that the library stands at the center of the campus. It also

occupies a corresponding place at the heart of the university's intellectual life. It pumps intellectual energy into every sector of the university, including laboratories where scientists download electronic journals and data bases without suspecting that they come from the library.

At Harvard, the university owes its name to its library, a collection of 400 books left to the College in 1638 by one John Harvard, and it grew up around that original core of learning. Today, 375 years and 18 million volumes later, we recognize that the university's greatness has derived above all from the intellectual riches built up over many generations and stored in the largest library system of any university in the world. But I don't mean to indulge in institutional bragging, which is an occupational disease to be avoided by librarians. Instead I want to ask a question: shouldn't Harvard's library and those of other research universities be considered as a national asset? Shouldn't their intellectual wealth be shared?

For most of history that wealth was restricted to a privileged few. And contrary to common belief, the history of libraries does not follow an uninterrupted, upward trajectory leading to the democratization of access to knowledge.

From what little we know about the ancient library of Alexandria, it functioned primarily to store texts, not to make them available to readers. It admitted a few scholars, but its main purpose probably was to embody the magnificence of the Ptolemaic dynasty. The same principle applied to the Chinese attempt to create a library that would contain all the books in the world – that is, the Chinese world. The Ching emperor Ch'ien-lung set out to do so by confiscating books from his subjects on a gigantic scale from 1772 to 1778. He kept everything that glorified the Ching dynasty and burned everything that was critical of it or favorable to the Mings – at least 2,320 books produced between 1550 and 1750. The Communist regime in Czechoslovakia used the country's libraries to purge literature rather than to preserve it. In 1954, it ordered all local librarians to cleanse their shelves of works that fit into categories that ranged from the expected ("fascism" and "pornography") to the bizarre ("formalism," "ruralism," and "snobbism".) They got rid of 7,500 works. And that was little compared with the 24 million

books destroyed in Soviet libraries during Stalin's Great Terror of 1938-39.

The history of libraries has a dark side, in the United States as well as other countries. The great African-American author Richard Wright could not borrow books from the public library in Memphis when he was a young man, because the color of his skin was black. The only way he could take out books was by pretending he was a servant fetching them for a white man.

To be sure, the oldest universities contributed greatly to civilization by building up their libraries. But they kept their books behind locked doors and thick walls, which removed them from outsiders. When I was a student at Oxford, the massive gate of my college slammed shut at 10:00 in the evening, and if you hadn't made it inside, you had to climb over one of the walls – a daunting experience, as the walls were ten to fifteen feet high and bristled at their top with spikes and shards of glass. A few secret passes existed, but even they were treacherous. In my favorite entry, you had to slip between rows of fixed and revolving spikes high up in the air.

But I want to discuss the invisible barriers to the knowledge stored in research libraries. Libraries frequently keep outsiders outside by all sorts of measures: restrictive qualifications for entry, payment to obtain a readers' card, and an atmosphere of intimidation. Ordinary folk hesitate to brave these barriers. They are kept at a distance by the learned elite, who wear an air of effortless superiority, which corresponds to the social sifting that the French sociologist Pierre Bourdieu identified as "distinction."

A counter-tendency gathered force in the age of Enlightenment, when philosophers like Condorcet understood the diffusion of knowledge to be the most important force in history, one that would extend everywhere, destroying prejudice and promoting progress. That faith was shared by the Founding Fathers of the United States, notably Condorcet's friend, Thomas Jefferson, who declared that "The field of knowledge is the common property of mankind." As the Founders understood it, the health of the republic depended on a well informed citizenry, and the spread of light was commensurate with the reading of books.

In retrospect, that vision looks utopian. But it helped inspire the opening up of libraries. Ordinary readers were allowed inside the

French Royal Library in 1692 and the British Museum in 1759. In the United States, the first large public library, established in Boston in 1848, allowed any citizen to borrow books and take them home to read. The New York Public Library opened its great collections in 1911 to anyone who walked in from the street. It served as an informal university for generations of immigrants who wanted both instruction and access to the literature in their native languages.

But this kind of access to knowledge depended on a limited technology: the printing press. Most people in the eighteenth century could not read, and most of those who were literate could not afford to buy books. Today we have the Internet. We now have it in our power to realize what was a utopian vision in the age of Enlightenment. As an example of what can now be accomplished, consider the difference between two great encyclopedias. The *Encyclopédie* edited by Diderot was the bible of the Enlightenment and a marvel of intellectual production 250 years ago: 17 volumes of text by about 200 contributors. But it cost 980 livres, the equivalent of two and a half years' income for a common laborer. Wikipedia now contains 30 million articles by 77,000 "active" contributors, and it reaches 365 million readers free of charge.

A new ideal of openness is transforming the world of knowledge. Its origins go back to the Enlightenment ideal of a republic of letters – that is, in principle, a free, intellectual realm with no police force, no boundaries, and no exclusiveness. But in practice, only a tiny elite enjoyed citizenship in this republic during the eighteenth century. Today we have open universities, open source software, open metadata, open access journals, and the beginnings of an open information highway.

Unfortunately, however, this tendency also has a darker side, because in some ways access to knowledge is being closed.

What, I ask you, is the cost of an average subscription to a chemical journal today? $4,044a year. What was it in 1970? $33a year. Inflation accounts for only a small part of the increase. During the last 25 years, the price of academic periodicals went up at four times the rate of the consumer price index. One year's subscription to the *Journal of Comparative Neurology* costs $29,113 – the equivalent of 600 monographs. Three giant publishers – Elsevier, Wiley-Blackwell, and Springer – now publish 42 percent of all academic

articles, and they make giant profits from them. In 2011, Elsevier turned a 36 percent profit on an income of 2 billion pounds sterling.

Don't get me wrong. I am not claiming that the ruinous increase in journal prices can be explained by greed on the part of the publishers. They are doing their job: returning the largest possible profit to their shareholders. Perhaps they should be congratulated for doing it so well, but their success should give us pause, because it illustrates the negative counterpart to the trend toward openness that I was just celebrating – namely commercialization, a trend toward *closing* access to knowledge.

The output of articles in medical science doubles every three to four years. So yes, more knowledge is constantly being produced, but an increasingly small proportion of it is available to the public. Why? Libraries can no longer afford to pay the prices. The average price for a medical journal was $12 in 1970; it now is $1,470. All over the country libraries are canceling subscriptions to academic journals, because they are caught between decreasing budgets and increasing costs. The logic of the bottom line is inescapable, but there is a *higher* logic that deserves consideration – namely, that the public should have access to knowledge produced with public funds.

Congress acted on that principle in 2008, when it required that articles based on grants from the National Institutes of Health be made available from an open-access repository, PubMed Central. But lobbyists for the publishers blunted that requirement by getting the NIH to accept a twelve-month embargo to prevent public accessibility long enough for them to cream off the demand. Not content with that victory, the lobbyists tried to abolish the NIH mandate in the so-called Research Works Act, a bill introduced in Congress in November 2011 and championed by Elsevier. The bill was withdrawn four months later following a wave of public protest, but the lobbyists are still at work, trying to block the Federal Research Public Access Act, which would give the public free access to the publication of all research funded by federal agencies with research budgets of $100 million or more. FASTR, as the bill is called, has been stalled in Congress, but its basic provisions have been adopted by a White House Directive of February 2013. In principle, therefore, the results of research funded by taxpayers will

be accessible to taxpayers and not monopolized by a few publishers. But what will be the practice? No one knows.

The battles over journal prices illustrate a conflict between two tendencies that will determine the digital future: democratization versus commercialization. How do those tendencies play out in the world of books and libraries?

The best example is provided by Google. In 2004 Google set out to digitize the collections of our greatest research libraries. It intended to use the data in a search service, which would make available snippets or short passages from the books for users seeking information about a specific subject. Authors and publishers sued Google for infringing their copyrights; and after four and a half years of strenuous negotiations, they reached a settlement, which transformed the search service into a gigantic commercial library. As a consequence, the research libraries that had originally supplied the books, free of charge, to Google would have to buy back access to the digital copies of the same books at a price that Google would determine. The cost they would have to pay for subscribing to Google's data base could escalate as disastrously as the price of subscriptions to academic journals, because the settlement had no provision for oversight or control by public authorities. In fact, the public was never consulted, and its interest was never considered. Fortunately, a New York federal court refused to authorize the settlement on the grounds that, among other things, it would create a monopoly in violation of the Sherman Anti-trust Act.

Google was actually proposing a monopoly of a new kind, a monopoly of access to information. It would have privatized a vast stretch of the public domain and collected a toll from anyone who tried to enter its fenced-off territory. It was an audacious and in some ways a thrilling project, and it raised the prospect of finding a democratic alternative to a commercial speculation. In the era of print, this open-access tradition goes back to the founding of the Boston Public Library in 1848. The inscription carved over its main entrance says, "Free to All." Now that we have entered the digital age, we can do better. We can make all of the material in all of our research libraries available to everyone free of charge. That is the basic idea behind the Digital Public Library of America, or DPLA.

On October 1, 2010, long before Google Book Search was rejected by the court, a group of leaders from foundations, libraries, and computer science met at Harvard to discuss the possibility of creating a Digital Public Library of America. The foundations would combine forces to provide the funds; the libraries would cooperate to furnish the books. We created a steering committee, a secretariat with a small staff, and work groups scattered across the country. Thousands of people participated, and everything was discussed openly – in large meetings in San Francisco, Chicago, and Washington; and also through email, list servs, web sites, wikis, and blogs. There was streaming, texting, tweeting – a running debate about every aspect of the plans in which everyone had a chance to be heard. At the same time, there was enough coordination from the Berkman Center at Harvard for the DPLA to be launched successfully online on April 18, 2013.

During the first week of its existence, the DPLA's website had 1.5 million downloads, and it registered 3.5 "pings" (brief consultations) every second. It offered 2.5 million items, free of charge, to the general public (28 percent of its users came from outside the United States), and that number doubled within nine months. The digital material was contributed by 1,100 institutions located in all 50 states.

How should you envisage the DPLA? Not as a grand edifice with an imposing dome erected over a gigantic data base, but rather as a "distributed" system – that is, a horizontal network that links up digital collections in libraries, archives, and museums in such a manner that users can get access to a document with one click on an electronic device. Now incorporated as a nonprofit enterprise, the DPLA has headquarters with a small staff in Boston, but it is not a top-down organization. Its horizontality corresponds to its democratic spirit and its basic goal – namely, to make the cultural heritage of America available, free of charge, to all Americans and in fact to everyone in the world.

Far from intending to serve an exclusive audience such as the elite who have access to university libraries, the DPLA is designed to meet the needs of many different publics – students of all ages, seniors in homes for the elderly, researchers without institutional affiliation, readers of every kind, including those who merely want to deepen their enjoyment of literature. It is organized in "hubs"

that extend its services like spokes of a wheel. "Content hubs" such as Harvard, the Smithsonian Institution, and the New York Public Library, provide digitized material from their enormous collections. "Service hubs" aggregate other collections and develop networks at the state level. They make special efforts to reach people in small communities, working with local public libraries. The libraries invite everyone in a town or an urban neighborhood to bring in diaries, letters, family photographs, and other items stored in attics or trunks. The material is digitized, given metadata, curated, and preserved. In this way communities sharpen their awareness of their own culture and history, and at the same time their local collections are integrated in a national network, which grows organically day by day.

Of course the DPLA cannot reach every town in the United States. It relies on volunteers, including a small army of "Community Reps" who stimulate activities within designated areas. Thanks to a grant from the Bill and Melinda Gates Foundation, a program is being developed to train librarians in public libraries to acquire the special skills that will be needed to launch digital projects in their communities.

The technological infrastructure of the DPLA has been designed to encourage local initiatives and "spin-off" projects of this kind. It, too, was a volunteer effort conducted on a national scale. For two and a half years, 1,100 computer scientists submitted ideas, which were combined and integrated by a team in the Berkman Center at Harvard University. When the system went live online (it can be consulted by anyone at dp.la), it functioned flawlessly. It is interoperable with Europeana, the system to integrate digital collections within the European Union, and it has special appeal because of the API (Application Programming Interface) built into its structure. The API provides a way for anyone to develop a particular tool or collection, which then can be used by everyone connected to the system. One such tool, "Book Shelf", creates the possibility of digital browsing. When consulting a particular book, users can see a row of related books whose spines appear on the computer screen as if they were physical books aligned on a shelf. One can click on the image of any spine, search through the book's table of contents and text, and construct an individualized collection of material on a certain subject.

In these and other "outreach" projects, the DPLA hopes to engage with its readers. Instead of simply making material available and waiting for them to use it, it seeks to interact with them and to enlist them in shaping its growth. It faces plenty of problems, of course. Having depended on support from foundations since 2010, it needs to develop a long-term business plan. Its technology requires constant maintenance and improvement. As it increases in size and scope, it must resolve issues of governance and administration.

Strange as it may seem, the biggest problem is legal. The DPLA must respect copyright, but copyright now covers books for the life of the author plus 70 years – that is, in most cases, for more than a century. Most literature from the twentieth century – everything after 1964 and most books published after 1923 – is therefore excluded from the DPLA's collection. There may be ways around this difficulty, especially if U.S. courts permit certain kinds of digital lending as "fair use." But for the moment, the DPLA is hoping that authors will voluntarily turn over the use of their copyrights after the commercial viability of their books has been exhausted. Once a book ceases to sell, most authors have one overriding desire: for their ideas to circulate through society rather than to remain unknown and out of reach to readers. The DPLA is working with an "Authors Alliance," to satisfy the needs of authors and readers by bringing them together digitally.

Despite the pressure of commercialization, therefore, the DPLA has tapped vein of public spirit that is both idealistic and pragmatic. It draws inspiration from the age of the Enlightenment, but it is designed to serve the needs of the twenty-first century; and at a time of disgust at the dysfunction of Washington, it has proved that it can get things done by independent initiative.

You should not think of it as a digital version of the Library of Congress. It's a new kind of library altogether, not just in its technology but also in its organization and spirit. It will operate simultaneously on many levels, personal, local, national, and international. It is already functioning successfully, and it will continue to acquire new functions, with ever-expanding collections, for many generations.

Technology will also continue to change, and the DPLA will have to change with it far into the future. But if we can get it right

now, we can help shape that future. For the first time in history, we can make the cultural heritage of humanity available to all humans. We have the technology, the know-how, the resources, and the will. We have taken the first steps, and now we have to get the job done.