

Digitization Wars, Redux

Karen Coyle

<http://kcoyle.net>, <https://kcoyle.blogspot.com>

March, 2021

From 2004 to 2016 the book world (authors, publishers, libraries, and booksellers) was involved in the complex and legally fraught activities around Google's book digitization project. Once known as "Google Book Search," the company claimed that it was digitizing books to be able to provide search services across the print corpus, much as it provides search capabilities over texts and other media that are hosted throughout the Internet.

Both the US Authors Guild and the Association of American Publishers sued Google (both separately and together) for violation of copyright. These suits took a number of turns including proposals for settlements that were arcane in their complexity and that ultimately failed. Finally, in 2016 the legal question was decided: digitizing to create an index is fair use as long as only minor portions of the original text are shown to users in the form of context-specific snippets.

We now have another question about book digitization: can books be digitized for the purpose of substituting remote lending in the place of the lending of a physical copy? This has been referred to as "Controlled Digital Lending (CDL)," a term developed by the Internet Archive for its online book lending services. The Archive has considerable experience with both digitization and providing online access to materials in various formats, and its Open Library site has been providing digital downloads of out of copyright books for more than a decade. Controlled digital lending applies solely to works that are presumed to be in copyright.

Controlled digital lending works like this: the Archive obtains and retains a physical copy of a book. The book is digitized and added to the Open Library catalog of works. Users can borrow the book for a limited time (2 weeks) after which the book "returns" to the Open Library. While the book is checked out to a user no other user can borrow that "copy." The digital copy is linked one-to-one with a physical copy, so if more than one copy of the physical book is owned then there is one digital loan available for each physical copy.

The Archive is not alone in experimenting with lending of digitized copies: some libraries have partnered with the Archive's digitization and lending service to provide digital lending for library-owned materials. In the case of the Archive the physical books are not available for lending. Physical libraries that are experimenting with CDL face the added step of making sure that the physical book is removed from circulation while the digitized book is on loan, and reversing that on return of the digital book.

Although CDL has an air of legality due to limiting lending to one user at a time, authors and publishers associations had raised objections to the practice. [nwu] However, in March of 2020 the Archive took a daring step that pushed their version of the CDL into litigation: using the closing of many physical libraries due to the COVID pandemic as its rationale, the Archive renamed its lending service the National Emergency Library [nel] and eliminated the one-to-one link between physical and digital copies. Ironically this meant that the Archive was then actually doing what the book industry had accused it of (either out of misunderstanding or as an exaggeration of the threat posed): it was making and lending digital copies beyond its physical holdings. The Archive stated that the National Emergency Library would last only until June of 2020, presumably because by then the COVID danger would have passed and libraries would have re-opened. In June the Archive's book lending service returned to the one-to-one model. Also in June a suit was filed by four publishers (Hachette, HarperCollins, Penguin Random House, and Wiley) in the US District Court of the Southern District of New York. [suit]

The Controlled Digital Lending, like the Google Books project, holds many interesting questions about the nature of "digital vs physical," not only in a legal sense but in a sense of what it means to read and to be a reader today. The lawsuit not only does not further our understanding of this fascinating question, it sinks immediately into hyperbole,

fear-mongering, and either mis-information or mis-direction. That is, admittedly, the nature of a lawsuit. What follows here is not that analysis but gives a few of the questions that are foremost in my mind.

Apples and Oranges

Each of the players in this drama has admirable reasons for their actions. The publishers explain in their suit that they are acting in support of authors, in particular to protect the income of authors so that they may continue to write. The Authors' Guild provides some data on author income, and by their estimate the average full-time author earns less than \$20,000 per year, putting them at poverty level.[aghard] (If that average includes the earnings of highly paid best selling authors, then the actual earnings of many authors is quite a bit less than that.)

The Internet Archive is motivated to provide democratic access to the content of books to anyone who needs or wants it. Even before the pandemic caused many libraries to close the collection housed at the Archive contained some works that are available only in a few research libraries. This is because many of the books were digitized during the Google Books project which digitized books from a small number of very large research libraries whose collections differ significantly from those of the public libraries available to most citizens.

Where the pronouncements of both parties fail is in making a false equivalence between some authors and all authors, and between some books and all books, and the result is that this is a lawsuit pitting apples against oranges. We saw in the lawsuits against Google that some academic authors, who may gain status based on their publications but very little if any income, did not see themselves as among those harmed by the book digitization project. Notably the authors in this current suit, as listed in the bibliography of pirated books in the appendix to the lawsuit, are ones whose works would be characterized best as “popular” and “commercial,” not academic: James Patterson, J. D. Salinger, Malcolm Gladwell, Toni Morrison, Laura Ingalls Wilder, and others. Not only do the living authors here earn above the poverty level, all of them provide significant revenue for the publishers themselves. And all of the books listed are in print and available in the marketplace. No mention is made of out-of-print books, no academic publishers seem to be involved.

On the part of the Archive, they state that their digitized books fill an educational purpose, and that their collection includes books that are not available in digital format from publishers:

“ While Overdrive, Hoopla, and other streaming services provide patrons access to latest best sellers and popular titles, the long tail of reading and research materials available deep within a library’s print collection are often not available through these large commercial services. What this means is that when libraries face closures in times of crisis, patrons are left with access to only a fraction of the materials that the library holds in its collection.”[cdl-blog]

This is undoubtedly true for some of the digitized books, but the main thesis of the lawsuit points out that the Archive has digitized and is also lending current popular titles. The list of books included in the appendix of the lawsuit shows that there are in-copyright and most likely in-print books of a popular reading nature that have been part of the CDL. These titles are available in print and may also be available as ebooks from the publishers. Thus while the publishers are arguing that current, popular books should not be digitized and loaned (apples), the Archive is arguing that they are providing access to items not available elsewhere, and for educational purposes (oranges).

The Law

The suit states that publishers are not questioning copyright law, only violations of the law.

“For the avoidance of doubt, this lawsuit is not about the occasional transmission of a title under appropriately limited circumstances, nor about anything permissioned or in the public domain. On the contrary, it is about IA’s purposeful collection of truckloads of in-copyright books to scan, reproduce, and then distribute digital bootleg versions online.” ([Suit] Page 3).

This brings up a whole range of legal issues in regard to distributing digital copies of copyrighted works. There have been lengthy arguments about whether copyright law could permit first sale rights for digital items, and the answer has generally been no; some copyright holders have made the argument that since transfer of a digital file is necessarily the making of a copy there can be no first sale rights for those files. [1stSale] [ag1] Some ebook systems, such as the Kindle, have allowed time-limited person-to-person lending for some ebooks. This is governed by license terms between Amazon and the publishers, not by the first sale rights of the analog world.

Section 108 of the copyright law does allow libraries and archives to make a limited number of copies. The first point of section 108 states that libraries can make a single copy of a work as long as 1) it is not for commercial advantage, 2) the collection is open to the public and 3) the reproduction includes the copyright notice from the original. This sounds to be what the Archive is doing. However, the next two sections (b and c) provide limitations on that first section that appear to put the Archive in legal jeopardy: section “b” clarifies that copies may be made for preservation or security; section “c” states that the copies can be made if the original item is deteriorating and a replacement can no longer be purchased. Neither of these applies to the Archive’s lending.

In addition to its lending program, the Archive provides downloads of scanned books in DAISY format for those who are certified as visually impaired by the National Library Service for the Blind and Physically Handicapped in the US. This is covered in 121A of the copyright law, Title 17, which allows the distribution of copyrighted works in accessible formats. This service could possibly be cited as a justification of the scanning of in-copyright works at the Archive, although without mitigating the complaints about lending those copies to others. This is a laudable service of the Archive if scans are usable by the visually impaired, but the DAISY-compatible files are based on the OCR’d text, which can be quite dirty. Without data on downloads under this program it is hard to know the extent to which this program benefits visually impaired readers.

Lending

Most likely as part of the strategy of the lawsuit, very little mention is made of “lending.” Instead the suit uses terms like “download” and “distribution” which imply that the user of the Archive’s service is given a permanent copy of the book

“With just a few clicks, any Internet-connected user can download complete digital copies of in-copyright books from Defendant.” ([suit] Page 2). “... distributing the resulting illegal bootleg copies for free over the Internet to individuals worldwide.” ([suit] Page 14).

Publishers were reluctant to allow the creation of ebooks for many years until they saw that DRM would protect the digital copies. It then was another couple of years before they could feel confident about lending - and by lending I mean lending by libraries. It appears that Overdrive, the main library lending platform for ebooks, worked closely with publishers to gain their trust. The lawsuit questions whether the lending technology created by the Archive can be trusted.

“...Plaintiffs have legitimate fears regarding the security of their works both as stored by IA on its servers” ([suit] Page 47).

In essence, the suit accuses IA of a lack of transparency about its lending operation. Of course, any collaboration between IA and publishers around the technology is not possible because the two are entirely at odds and the publishers would reasonably not cooperate with folks they see as engaged in piracy of their property.

Even if the Archive’s lending technology were proven to be secure, lending alone is not the issue: the Archive copied the publishers’ books without permission prior to lending. In other words, they were lending content that they neither owned (in digital form) nor had licensed for digital distribution. Libraries pay, and pay dearly, for the ebook lending service that they provide to their users. The restrictions on ebooks may seem to be a money-grab on the part of publishers, but from their point of view it is a revenue stream that CDL threatens.

Is it About the Money?

“... IA rakes in money from its infringing services...” ([suit] Page 40). (Note: publishers earn, IA “rakes in”)

“Moreover, while Defendant promotes its non-profit status, it is in fact a highly commercial enterprise with millions of dollars of annual revenues, including financial schemes that provide funding for IA’s infringing activities. ([suit] Page 4).

These arguments directly address section (a)(1) of Title 17, section 108: “(1) the reproduction or distribution is made without any purpose of direct or indirect commercial advantage”.

At various points in the suit there are references to the Archive’s income, both for its scanning services and donations, as well as an unveiled show of envy at the over \$100 million that Brewster Kahle and his wife have in their jointly owned foundation. This is an attempt to show that the Archive derives “direct or indirect commercial advantage” from CDL. Non-profit organizations do indeed have income, otherwise they could not function, and “non-profit” does not mean a lack of a revenue stream, it means returning revenue to the organization instead of taking it as profit. The argument relating to income is weakened by the fact that the Archive is not charging for the books it lends. However, much depends on how the courts will interpret “indirect commercial advantage.” The suit argues that the Archive benefits generally from the scanned books because this enhances the Archive’s reputation which possibly results in more donations. There is a section in the suit relating to the “sponsor a book” program where someone can donate a specific amount to the Archive to digitize a book. How many of us have not gotten a solicitation from a non-profit that makes statements like: “\$10 will feed a child for a day; \$100 will buy seed for a farmer, etc.”? The attempt to correlate free use of materials with income may be hard to prove.

Reading

Decades ago, when the service Questia was just being launched (Questia ceased operation December 21, 2020), Questia sales people assured a group of us that their books were for “research, not reading.” Google used a similar argument to support its scanning operation, something like “search, not reading.” The court decision in Google’s case decided that Google’s scanning was fair use (and transformative) because the books were not available for reading, as Google was not presenting the full text of the book to its audience.[suit-g]

The Archive has taken the opposite approach, a “books are for reading” view. Beginning with public domain books, many from the Google books project, and then with in-copyright books, the Archive has promoted reading. It developed its own in-browser reading software to facilitate reading of the books online. [reader]

Although the publishers sued Google for its scanning, they lost due to the “search, not reading” aspect of that project. The Archive has been very clear about its support of reading, which takes the Google justification off the table.

“Moreover, IA’s massive book digitization business has no new purpose that is fundamentally different than that of the Publishers: both distribute entire books for reading.” ([suit] Page 5).

However, the Archive's statistics on loaned books shows that a large proportion of the books are used for 30 minutes or less.

“Patrons may be using the checked-out book for fact checking or research, but we suspect a large number of people are browsing the book in a way similar to browsing library shelves.” [ia1]

In its article on the CDL, the Center for Democracy and Technology notes that “the majority of books borrowed through NEL were used for less than 30 minutes, suggesting that CDL’s primary use is for fact-checking and

research, a purpose that courts deem favorable in a finding of fair use.” [cdt] The complication is that the same service seems to be used both for reading of entire books and as a place to browse or to check individual facts (the facts themselves cannot be copyrighted). These may involve different sets of books, once again making it difficult to characterize the entire set of digitized books under a single legal claim.

The publishers claim that the Archive is competing with them using pirated versions of their own products. That leads us to the question of whether the Archive’s books, presented for reading, are effectively substitutes for those of the publishers. Although the Archive offers actual copies, those copies that are significantly inferior to the original. However, the question of quality did not change the judgment in the lawsuit against copying of texts by Kinko’s [kinkos], which produced mediocre photocopies from printed and bound publications. It seems unlikely that the quality differential will serve to absolve the Archive from copyright infringement even though the poor quality of some of the books interferes with their readability.

Digital is Different

Publishers have found a way to monetize digital versions, in spite of some risks, by taking advantage of the ability to control digital files with technology and by licensing, not selling, those files to individuals and to libraries. It’s a “new product” that gets around First Sale because, as it is argued, every transfer of a digital file makes a copy, and it is the making of copies that is covered by copyright law. [1stSale]

The upshot of this is that because a digital resource is licensed, not sold, the right to pass along, lend, or re-sell a copy (as per Title 17 section 109) does not apply even though technology solutions that would delete the sender’s copy as the file safely reaches the recipient are not only plausible but have been developed. [resale]

“Like other copyright sectors that license education technology or entertainment software, publishers either license ebooks to consumers or sell them pursuant to special agreements or terms.” ([suit] Page 15)

“When an ebook customer obtains access to the title in a digital format, there are set terms that determine what the user can or cannot do with the underlying file.”([suit] Page 16)

This control goes beyond the copyright holder’s rights in law: DRM can exercise controls over the actual use of a file, limiting it to specific formats or devices, allowing or not allowing text-to-speech capabilities, even limiting copying to the clipboard.

Publishers and Libraries

The suit claims that publishers and libraries have reached an agreement, an equilibrium.

“To Plaintiffs, libraries are not just customers but allies in a shared mission to make books available to those who have a desire to read, including, especially, those who lack the financial means to purchase their own copies.” ([suit] Page 17).

In the suit, publishers contrast the Archive’s operation with the relationship that publishers have with libraries. In contrast with the Archive’s lending program, libraries are the “good guys.”

“... the Publishers have established independent and distinct distribution models for ebooks, including a market for lending ebooks through libraries, which are governed by different terms and expectations than print books.”([suit] Page 6).

These “different terms” include charging much higher prices to libraries for ebooks, limiting the number of times an ebook can be loaned. [pricing1] [pricing2]

“Legitimate libraries, on the other hand, license ebooks from publishers for limited periods of time or a limited number of loans; or at much higher prices than the ebooks available for individual purchase.” [agol]

The equilibrium of which publishers speak looks less equal from the library side of the equation: library literature is replete with stories about the avarice of publishers in relation to library lending of ebooks. Some authors/publishers even speak out against library lending of ebooks, claiming that this cuts into sales. (This same argument has been made for physical books.)

“If, as Macmillan has determined, 45% of ebook reads are occurring through libraries and that percentage is only growing, it means that we are training readers to read ebooks for free through libraries instead of buying them. With author earnings down to new lows, we cannot tolerate ever-decreasing book sales that result in even lower author earnings.” [agliblend][ag42]

The ease of access to digital books has become a boon for book sales, and ebook sales are now rising while hard copy sales fall. This economic factor is a motivator for any of those engaged with the book market. The Archive’s CDL is a direct affront to the revenue stream that publishers have carved out for specific digital products. There are indications that the ease of borrowing of ebooks - not even needing to go to the physical library to borrow a book - is seen as a threat by publishers. This has already played out in other media, from music to movies.

It would be hard to argue that access to the Archive’s digitized books is merely a substitute for library access. Many people do not have actual physical library access to the books that the Archive lends, especially those digitized from the collections of academic libraries. This is particularly true when you consider that the Archive’s materials are available to anyone in the world with access to the Internet. If you don’t have an economic interest in book sales, and especially if you are an educator or researcher, this expanded access could feel long overdue.

We need numbers

We really do not know much about the uses of the Archive’s book collection. The lawsuit cites some statistics of “views” to show that the infringement has taken place, but the page in question does not explain what is meant by a “view”. Archive pages for downloadable files of metadata records also report “views” which most likely reflect views of that web page, since there is nothing viewable other than the page itself. Open Library book pages give “currently reading” and “have read” stats, but these are tags that users can manually add to the page for the work. To compound things, the 127 books cited in the suit have been removed from the lending service (and are identified in the Archive as being in the collection “litigation works”).

Although numbers may not affect the legality of the controlled digital lending, the social impact of the Archive’s contribution to reading and research would be clearer if we had this information. Although the Archive has provided a small number of testimonials, a proof of use in educational settings would bolster the claims of social benefit which in turn could strengthen a fair use defense.

Citations

[1stSale] <https://abovethelaw.com/2017/11/a-digital-take-on-the-first-sale-doctrine/>

[ag1] <https://www.authorsguild.org/industry-advocacy/reselling-a-digital-file-infringes-copyright/>

[ag42] <https://www.authorsguild.org/industry-advocacy/authors-guild-survey-shows-dramatic-42-percent-decline-in-authors-earnings-in-last-decade/>

[aghard] <https://www.authorsguild.org/the-writing-life/why-is-it-so-goddamned-hard-to-make-a-living-as-a-writer-today/>

[aglibend] <https://www.authorsguild.org/industry-advocacy/macmillan-announces-new-library-lending-terms-for-ebooks/>

[agol] <https://www.authorsguild.org/industry-advocacy/update-open-library/>

[cdl-blog] <https://blog.archive.org/2020/03/09/controlled-digital-lending-and-open-libraries-helping-libraries-and-readers-in-times-of-crisis/>

[cdt] <https://cdt.org/insights/up-next-controlled-digital-lendings-first-legal-battle-as-publishers-take-on-the-internet-archive/>

[kinkos] <https://law.justia.com/cases/federal/district-courts/FSupp/758/1522/1809457>

[nel] <http://blog.archive.org/national-emergency-library/>

[nwu] "[Appeal from the victims of Controlled Digital Lending \(CDL\)](#)". (Retrieved 2021-01-10)

[pricing1] <https://www.authorsguild.org/industry-advocacy/e-book-library-pricing-the-game-changes-again/>

[pricing2] <https://americanlibrariesmagazine.org/blogs/e-content/ebook-pricing-wars-publishers-perspective/>

[reader] [Bookreader](#)

[resale] <https://www.hollywoodreporter.com/thr-esq/appeals-court-weighs-resale-digital-files-1168577>

[suit] <https://www.courtlistener.com/recap/gov.uscourts.nysd.537900/gov.uscourts.nysd.537900.1.0.pdf>

[suit-g] <https://cases.justia.com/federal/appellate-courts/ca2/13-4829/13-4829-2015-10-16.pdf?ts=1445005805>