

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

-----	x
HACHETTE BOOK GROUP, INC.,	:
HARPERCOLLINS PUBLISHERS LLC, JOHN	:
WILEY & SONS, INC., and PENGUIN RANDOM	: 20 Civ. _____
HOUSE LLC,	:
	: ECF Case
Plaintiffs,	:
	: COMPLAINT
-against-	: TRIAL BY JURY DEMANDED
	:
INTERNET ARCHIVE and DOES 1 through 5,	:
inclusive,	:
Defendants.	:
-----	x

Plaintiffs Hachette Book Group, Inc. (“Hachette”), HarperCollins Publishers LLC (“HarperCollins”), John Wiley & Sons, Inc. (“Wiley”), and Penguin Random House LLC (“Penguin Random House”), by and through their attorneys Davis Wright Tremaine LLP and Oppenheim + Zebtrak, LLP, for their Complaint, hereby allege against Defendant Internet Archive (“IA” or “Defendant”) and Does 1 through 5 as follows:

NATURE OF THE ACTION

1. Plaintiffs Hachette, HarperCollins, Penguin Random House, and Wiley (collectively, “Plaintiffs” or “Publishers”) bring this copyright infringement action against IA in connection with website operations it markets to the public as “Open Library” and/or “National Emergency Library.” Plaintiffs are four of the world’s preeminent publishing houses. Collectively, they publish some of the most successful and leading authors in the world, investing in a wide range of fiction and nonfiction books for the benefit of readers everywhere. All of the Plaintiffs are member companies of the Association of American Publishers, the mission of which is to be the voice of American publishing on matters of law and public policy.

2. Defendant IA is engaged in willful mass copyright infringement. Without any license or any payment to authors or publishers, IA scans print books, uploads these illegally scanned books to its servers, and distributes verbatim digital copies of the books *in whole* via public-facing websites. With just a few clicks, any Internet-connected user can download *complete* digital copies of in-copyright books from Defendant.

3. The scale of IA's scheme is astonishing: At its "Open Library," located at www.openlibrary.org and www.archive.org (together, the "Website"), IA currently distributes digital scanned copies of over *1.3 million* books. And its stated goal is to do so for millions more, essentially distributing free digital copies of every book ever written. Despite the "Open Library" moniker, IA's actions grossly exceed legitimate library services, do violence to the Copyright Act, and constitute willful digital piracy on an industrial scale. Consistent with the deplorable nature of piracy, IA's infringement is intentional and systematic: it produces mirror-image copies of millions of unaltered in-copyright works for which it has no rights and distributes them in their entirety for reading purposes to the public for free, including voluminous numbers of books that are currently commercially available.

4. Books have long been essential to our society. Fiction and non-fiction alike, they transport us to new worlds, broaden our horizons, provide us with perspective, reflect the ever-growing knowledge of humanity in every field, spark our imaginations and deepen our understanding of the world. Yet, books are not self-generating. They are the product of training and study, talent and grit, perseverance and creativity, investment and risk, and untold hours of work.

5. The publishing ecosystem not only depends upon copyright law, it is historically intertwined with the founding of the United States. In 1787, the Framers adopted the Copyright

Clause of the Constitution, explicitly authorizing Congress “[t]o 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.” U.S. Const., Art. I, §8, cl. 8. In 1790, the First Congress enacted the first Copyright Act, focused on incentivizing both the creation *and legal dissemination* of books, maps, and charts. Congresses ever since have carefully balanced copyright amendments to advance the public good and for more than 200 years have prescribed to authors a suite of enforceable exclusive rights to their writings—which publishers, in turn, encourage, invest in, license, and distribute to readers through bookstores, libraries, and a multitude of e-commerce platforms. In this process of publishing books that educate, entertain, and inspire the public, publishers rely not only on the exclusive rights that are their lifeblood, but on the expectation that Congress has carefully considered and appropriately tailored any limitations and exceptions to said rights.

6. IA not only acts entirely outside any legal framework, it does so flagrantly and fraudulently. And it proceeds despite actual notice that its actions constitute infringement. 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. IA’s Website includes books of every stripe—from bestsellers to scholarly monographs, from entertaining thrillers and romances to literary fiction, from self-help books to biographies, from children’s books to adult books. IA often suggests that the Website is limited to twentieth-century books, but this is neither accurate nor a defense. IA scans, uploads, and distributes huge numbers of in-copyright books published in both the twentieth and twenty-first centuries, including many books

published within just the past few years. IA's unauthorized copying and distribution of Plaintiffs' works include titles that the Publishers are currently selling commercially and currently providing to libraries in ebook form, making Defendant's business a direct substitute for established markets. Free is an insurmountable competitor.

7. Publishers have long supported public libraries, recognizing the significant benefits to the public of ready access to books and other publications. This partnership turns upon a well-developed and longstanding library market, through which public libraries buy print books and license ebooks (or agree to terms of sale for ebooks) from publishers, usually via book wholesalers or library ebook aggregators. IA's activities are nothing like those of public libraries, but rather the kind of quintessential infringement that the Copyright Act directly prohibits. 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. By branding itself with the name "Open Library," it thus badly misleads the public and boldly misappropriates the goodwill that libraries enjoy and have legitimately earned.

8. IA defends its willful mass infringement by asserting an invented theory called "Controlled Digital Lending" ("CDL")—the rules of which have been concocted from whole cloth and continue to get worse. For example, at first, under this theory IA claimed to limit the number of scanned copies of a title available for free download at any one time to the number of print books of that title in its collection—though no provision under copyright law offers a colorable defense to the systematic copying and distribution of digital book files simply because the actor collects corresponding physical copies. Then, in the face of the COVID-19 pandemic, IA opportunistically seized upon the global health crisis to further enlarge its cause, announcing

with great fanfare that it would remove these already deficient limitations that were purportedly in place. Today, IA offers an enormous universe of scanned books to an unlimited number of individuals simultaneously in its “National Emergency Library.” IA’s blatant, willful infringement is all the more egregious for its timing, which comes at the very moment that many authors, publishers, and independent bookstores, not to mention libraries, are both struggling to survive amidst economic uncertainty and planning deliberately for future, changing markets.

9. Under whatever guise IA attempts to frame its massive infringement—whether adopting the invented CDL theory or filling the self-appointed role as “National Emergency Library”—its actions find no support in the Copyright Act. IA’s defenses of its actions—both before and after the onset of the COVID-19 crisis—are baseless. First, while IA claims to serve an educational purpose, education has long been a primary mission and market of publishers. It is authors and publishers who create the books of scholarship and literature for educators, students, and other readers; IA creates nothing. IA plays no role in the hard work of researching, writing, or publishing the works or, for that matter, in creating or sustaining the overall publishing ecosystem and its distinct partnerships and markets. Nor does IA contribute to the underlying scholarship through commentary or criticism. 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. In short, Defendant merely exploits the investments that publishers have made in their books, and it does so through a business model that is designed to free-ride on the work of others. Defendant pays for none of the expenses that go into publishing a book and is nothing more than a mass copier and distributor of bootleg works. In so doing, IA undermines the balance and promise of copyright law by usurping the

Publishers' ability to license and sell the books that they have lawfully produced on behalf of authors and for the benefit of readers.

10. IA's self-serving assertion and promotion of "Controlled Digital Lending" as both an actual legal doctrine and a justification for its infringement affronts the most basic realities of the law and the markets it propels. As a matter of markets, IA's one-to-one conflation of print and ebooks is fundamentally flawed. Digital books are inherently different from physical books. They can fly around the world in a second; they do not degrade over time as physical books do; and they require devices to read them. For these reasons, 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. IA's end-run around these differences and restrictions is aggressive and unlawful. In short, all of the reasons why IA has scanned print books to create digital files are the very same reasons why authors and publishers provide digital books under different terms than print books—as they are entitled to do under the Copyright Act.

11. No concept of fair use supports the systematic mass copying or distribution of entire books for the purpose of mass reading, or put another way, for the purpose of providing to readers the very thing that publishers and authors provide in the first place through lawful and established channels. IA does not add something new to the Plaintiffs' books, with a different purpose or character; thus, it cannot even begin to make the all-important showing that its use of the works is transformative. Separately, Section 109 of the Copyright Act is clear that, pursuant to the doctrine of first sale, the owner of a lawfully acquired print book may dispose only of her/his particular print copy. One who makes and distributes reproductions of that physical copy—such as IA's low quality scans—is well outside the bounds of the law.

12. Nor do IA’s efforts to brand itself as a library somehow imbue it with any right to digitize and distribute unauthorized digital copies of books. Libraries are trusted institutions that serve the communities that fund them. When Congress contemplated the making of digital copies by libraries under 17 U.S.C. §108, it engaged all relevant stakeholders and created a set of rational, targeted exceptions to infringement liability—exceptions that have no application to IA’s actions. As the Copyright Office observed in a relevant public study titled “Legal Issues in Mass Digitization” (October 2011), “The Section 108 exception does not contemplate mass digitization.”

13. The creation, publication, and distribution of books is an ecosystem. IA disaggregates itself from this ecosystem, ignores the law, and asserts that its goal of providing free copies of books somehow excuses it from any responsibility to those who have created the works and hold exclusive rights under the Copyright Act. Its goal of creating digital copies of books and providing them to whomever wants to download them reflects a profound misunderstanding of the costs of creating books, a profound lack of respect for the many contributors involved in the publication process, and a profound disregard of the boundaries and balance of core copyright principles. IA does not seek to “free knowledge”; it seeks to destroy the carefully calibrated ecosystem that makes books possible in the first place—and to undermine the copyright law that stands in its way.

14. In sum, IA’s massive taking violates the Plaintiffs’ exclusive rights under 17 U.S.C §106. Plaintiffs bring this action to halt IA’s assault on their rights.

THE PARTIES

A. Plaintiffs

15. Plaintiffs are four of the leading book publishers in the United States. Working closely with their authors, Plaintiffs source, develop, edit, publish, market, and distribute tens of

thousands of books per year, across the full spectrum of genres and topics.

16. Plaintiff Hachette is a publishing company, organized under the laws of Delaware, with its principal place of business at 1290 Sixth Avenue, New York, NY 10104. With a history stretching back to 1837, Hachette works with bestselling authors who have been published all over the world. Hachette books and authors have won Pulitzer Prizes, National Book Awards, Newbery Medals, Caldecott Medals, and Nobel Prizes. Its many publishing imprints include prominent brands such as Little, Brown and Company, Little, Brown Books for Young Readers, Grand Central Publishing, Basic Books, Public Affairs, Orbit, FaithWords and Center Street.

17. Plaintiff HarperCollins is a publishing company, organized under the laws of Delaware, with its principal place of business at 195 Broadway, New York, NY 10007. HarperCollins has more than 200 years of history in the book publishing industry and the company now operates more than 120 imprints and brands in 17 countries worldwide. Each year, HarperCollins publishes approximately 10,000 new books in more than a dozen languages and boasts a catalogue of more than 200,000 titles in print and digital formats. Working across a wide range of genres, authors published by HarperCollins have won the Nobel Prize, the Pulitzer Prize, the National Book Award, the Newbery and Caldecott Medals, and the Man Booker Prize, among other honors.

18. Plaintiff Penguin Random House is a publishing company, organized under the laws of Delaware, with its principal place of business at 1745 Broadway, New York, NY 10019. Penguin Random House can trace its history back to the mid-nineteenth century and one of its progenitors, Random House, published the first authorized edition of James Joyce's *Ulysses* in the English-speaking world, among other landmark titles. The portfolio operated by Penguin

Random House has grown to encompass nearly 275 independent imprints and brands across five continents. Penguin Random House publishes 15,000 new titles per year—catering to readers of all ages and at every stage of life—and sells close to 800 million print books, audiobooks, and ebooks annually. It has published hundreds of the most widely read authors in the world.

19. Plaintiff Wiley is a publishing company, organized under the laws of New York, with its principal place of business at 111 River Street, Hoboken, NJ 07030. Founded in 1807, Wiley has over 200 years of experience publishing scientific, professional, and education books and journals in print and digital formats. Wiley has published works by over 450 Nobel Laureates. It publishes over 2,000 new books each year and currently offers over 120,000 titles.

20. Plaintiffs are the copyright owners or owners of exclusive rights under copyright in, *inter alia*, each of the works listed in Exhibit A (the “Works”), on which they bring suit here. Exhibit A is an illustrative, non-exhaustive list of in-copyright works that Defendant(s) infringed through the activities complained of herein. Upon information and belief, all of the Works have been scanned and uploaded to the Website by IA. All of these titles are commercially available.

21. The Works represent a cross-section of the exceptional books that are made possible by a functioning publishing ecosystem, from perennial classic novels to more recent highly acclaimed works of non-fiction and everything in between. Some of the greatest works of fiction ever published find their place among the Works in suit, including *The Lord of the Flies* by William Golding (winner of the 1983 Nobel Prize for Literature), *Song of Solomon* by Toni Morrison (winner of the 1993 Nobel Prize for Literature), and *Their Eyes Were Watching God* by Zora Neal Hurston. The Works also include *New York Times* bestselling authors like John Grisham and James Patterson and equally popular thrillers like *Gone Girl* by Gillian Flynn as well as hard-hitting contemporary novels, such as *The Miseducation of Cameron Post* by Emily

M. Danforth and *The Road* by Cormac McCarthy, which won the 2007 Pulitzer Prize for Fiction. Children's books are well-represented, from old favorites like *Little House on the Prairie* and *The Lion, The Witch and the Wardrobe* to more recent treasures, like the works of Lemony Snicket and *Escape from Mr. Lemoncello's Library*. Books for young adults are also included, like *Scat* by Carl Hiaasen and *The House on Mango Street* by Sandra Cisneros.

22. No less important than the works of fiction are the outstanding examples of non-fiction books represented by the Works. The Works contain multiple titles from the ever-popular "For Dummies" series, which have taught intrepid readers the basics of everything from oil painting to comparative religion. For those looking for success in business, the Works include books by the management guru Patrick Lencioni and books on investment by billionaire analyst Ken Fisher. Also included are Malcolm Gladwell's highly influential works on psychology and behavioral economics, a work by Nobel Peace Prize winner Elie Wiesel, and the well-loved *A Short History of Nearly Everything* by Bill Bryson.

B. Defendant

23. Defendant IA is a 501(c)(3) corporation, organized under the laws of California, with its principal place of business at 300 Funston Avenue, San Francisco, CA 94118.

24. IA is registered with the New York Department of State to transact business and accept service of process within the State of New York. IA currently transacts business within the State of New York and this District by, *inter alia*, distributing digital copies of books (and other content) to New York residents over the Internet, by providing New York residents with services-for-a-fee related to the digitization of books, and by soliciting and accepting contributions from New York residents to further its digitization and distribution of books. In addition, certain Works were copied and digitized by IA in New York.

25. IA harms the Publishers in this District because IA has copied and uploaded the Publishers' copyrighted books to its Website, including each of the Works in suit, without permission, and IA currently distributes copies to users of the Website, in New York or elsewhere. Upon information and belief, many of the acts of copyright infringement committed by IA set forth in this Complaint occurred within this State and District—including illegal reproductions, distributions, public displays, and/or public performances. Both the Publishers' economic and author relations damages are primarily felt in this State, where three of the Plaintiffs have their principal place of business and the fourth (Wiley) is incorporated. IA knew it would cause injury to Publishers in this State and District, or it should have reasonably expected injury to occur here. Indeed, IA acknowledges that in the last thirty days over 151,000 views on its site came from New York State, making New York the jurisdiction with the third highest IA views in the world. *See* Internet Archive, Books to Borrow, <https://archive.org/details/inlibrary?tab=about> (last accessed May 29, 2020).

26. IA derives substantial revenue from interstate and international commerce. According to public filings, IA has earned over \$100 million in the last ten years from a national network of supporters, at least some of whom are based in New York, and from the services it sells to clients in New York and all over the United States, including industrial-scale book scanning services.

27. Defendants Doe 1 through Doe 5 are certain individuals or entities whose true identities are not currently known to Plaintiffs. Defendants Doe 1 through Doe 5, who are sued under fictitious names, are those who also may be responsible for the unlawful activities complained of herein. (Doe 1 through Doe 5 do not include any public, university, or academic

libraries.) Once Plaintiffs ascertain their identities, Plaintiffs will seek leave of the Court to amend the Complaint to include Defendants Doe 1 through Doe 5 as named defendants.

JURISDICTION AND VENUE

28. This Court has subject matter jurisdiction over this action, which arises under the Copyright Act of 1976, 17 U.S.C. §§ 101, *et seq.*, pursuant to 28 U.S.C. §§ 1331 and 1338.

29. This Court has personal jurisdiction over Defendant pursuant to CPLR 302 because, *inter alia*, Defendant transacts business within the State of New York and supplies services in this State; because Defendant has committed tortious acts within the State of New York, including the direct and indirect infringement of Plaintiffs' exclusive rights in copyrighted books; and, because Defendant has caused injury to Plaintiffs in this State by allowing Internet users to download and view Plaintiffs' Works for free on the infringing Website and knew or reasonably should have known its acts would have consequences in this State, all while deriving millions of dollars in revenue from interstate commerce,.

30. This Court independently has personal jurisdiction over Defendant pursuant to CPLR 301 because IA is registered to do business in the State of New York and has pervasive corporate ties to the State that are sufficient to justify the imposition of general jurisdiction here.

31. Venue is proper in this District under 28 U.S.C. §§ 1391 and 1400.

FACTUAL BACKGROUND

A. The Book Publishing Ecosystem

i. Publishers and Authors Rely on Copyright Law to Create Functioning Markets for Books

32. Books are a cornerstone of our culture and system of democratic self-government and play a critical role in education. Because books require so much time and effort to write and develop, they offer the promise of high-quality expression, important insights, and long-term

value to society. But the qualities that make books among the most reliable conduits of learning and most intensive sources of creative expression come at a high cost. Publishers are largely responsible for bearing these costs and, in doing so, act as leading defenders of free speech, promoters of literacy and scientific knowledge, and creators of the stories people thrive on.

33. Authors devote great effort and care in researching and writing books, a skill which requires training and imagination. It is not unusual for an author to spend years writing a single book of fiction or nonfiction. Many authors, from the most celebrated *New York Times* bestsellers to new talents still making a name for themselves, write books for a living and rely on income from writing as their primary means of support. Writing is an expert craft, but one that is commercially unpredictable. Authors who find success with one book may benefit from renewed interest in and sales of their previous titles. Copyright law supports this long potential by ensuring a lengthy term of protection, as well as licensing and sale possibilities.

34. In the United States, publishing dates back to the dawn of our democracy. Over hundreds of years, book publishers like Plaintiffs have invested in the talent of authors and developed unparalleled expertise in the art of publishing high-quality books by providing a variety of vital services including editing, marketing, and distribution. Publishers expend the necessary resources, financial and otherwise, in reliance on the enforceable exclusive rights afforded by copyright law that make recouping expenditures possible. The steadfastness of the law, in turn, promotes new technologies and new business models and distribution mechanisms by which to reach new audiences, no matter the circumstances. Indeed, at a time of crisis such as we have now with COVID-19, the continued viability of publishing is more important than ever to society.

35. The founders of this nation wrote a copyright clause into the Constitution to empower Congress to incentivize authors to create and publish their work, yielding a robust history of Copyright Acts that date to 1790 and have always squarely addressed the protection of books. The Copyright Act of 1976—which enacts the constitutional imperative into law and balances the rights of readers with the rights of copyright owners—enables authors to profit from writing books by granting them the legal right to control the reproduction, distribution, public display, and public performance of their work, and to create derivative works, among other exclusive rights. Each of these rights is implicated in this action.

36. In a related fashion, copyright law gives authors and publishers, as rightsholders, exclusive control over how to publish their content in order to allow book markets to develop and thrive. This includes empowering publishers to tailor their means of distribution and terms of sale or license depending on the format or medium in which a particular title is released. These carefully calibrated markets are precisely the markets that IA seeks to disrupt and destroy by arrogating to itself the right to engage in bulk digitization of the Publishers' in-copyright books without a license and without any compensation, and by distributing the resulting illegal bootleg copies for free over the Internet to individuals worldwide.

37. Over hundreds of years publishers have found ways to maintain viable markets for books even as revolutions in publishing have driven changes in format, from leather-bound hardcover books to paperbacks to the paperless ebooks we read on digital devices. The ability of publishers to develop a diversity of new channels as technology evolves is crucial to meeting the high cost of publishing.

38. This includes the market for both new and “backlist” books. Book publishers derive substantial revenue from backlist books, which range from venerable classics like *The*

Bell Jar by Sylvia Plath or *Winds of War* by Herman Wouk to works written only a few years ago, including bestselling works such as *Eat, Pray, Love* by Elizabeth Gilbert and *Commonwealth* by Ann Patchett, all of which are Works in suit. Moreover, a great deal of the most successful children's books, including many of the Works, are backlist books.

ii. The Development of Functioning Markets for Ebooks

39. The rise of a commercial market for ebooks provides an example of publishers' adapting to new technologies to create new and diverse channels to meet the demands of readers. Since the early 2000s, Plaintiffs and other publishers have offered readers digital versions of their books, which can be read on portable electronic devices such as the Kindle, Nook, iPad, and other smart devices. Since that time, ebooks have grown to become a major source of revenue for authors and publishers. Publishers have invested heavily to expand the ebook market, including by publishing their backlist titles in ebook form. They have devoted considerable time, money, and professional expertise to create high-quality ebooks to deliver to readers. Authors rely on publishers to present their works well to readers.

40. The fundamental differences between print and digital formats require publishers to market print books and ebooks in different ways. Not only are the cost structures and distribution systems different for these two formats, but ebooks are digital files that can pose significant security concerns. Without protective measures, digital files can be copied perfectly, instantaneously and in practically infinite quantity at virtually no cost, and distributed all over the world in a split second.

41. Because of these material differences in format, publishers do not distribute ebooks the same way that they sell traditional paper books. 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 established by each publisher

and the platforms on which the ebooks may be read. By contrast, they sell copies of print books without any restrictions.

42. 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. Publishers also use digital rights management technology (“DRM”) to restrict the use and further distribution of ebook files. The commercial ebook market would not be viable if publishers lacked the ability to place any control over the means of distribution of ebooks or to prevent unlimited copying or distribution of the files.

43. Copyright law recognizes and enforces the right of copyright owners to control their works through DRM technology. In 1998, Congress enacted the Digital Millennium Copyright Act (“DMCA”), which made it illegal to circumvent DRM technology. The rights that enable publishers to control the publication of ebooks ultimately benefit readers because they enable publishers to provide their ebooks to a variety of channels and at a variety of price points that make sense for the specific authors and titles in their catalog.

44. Over the past twenty years, publishers have embraced the shift to digital commerce and correspondingly offered consumers an ever-evolving variety of formats, distribution, and access models. Indeed, Plaintiffs have collectively made thousands of the books in their catalogs available in ebook form, including the Works identified in this suit and hundreds of thousands of other works, backlist titles included. Virtually all of the trade books being offered for sale by Plaintiffs, including backlist books, are available in both print and ebook form.

45. At the same time, it is a basic tenet of copyright law that the copyright holder retains the exclusive right to decide whether or not to publish a copyrighted book in a digital

format. At times, authors and publishers make the purposeful decision not to publish particular titles as ebooks. Some niche categories of books are excluded from the ebook market because they are not suitable for that market. In other cases, authors and publishers decide not to publish ebooks because a digital edition would be far inferior to a print version—as with some heavily illustrated books, art books, or creatively designed children’s books.

iii. The Established Market Equilibrium Between Authors, Publishers, and Libraries

46. Public libraries are among the most cherished institutions in this country. 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. Much like a publisher must decide which books to invest in—balancing all sorts of considerations—librarians must make reasoned decisions about which books to purchase in order to best serve the needs of their communities.

47. Plaintiffs have worked with libraries and library aggregators such as OverDrive to pioneer an innovative and highly successful service that enables library patrons to easily and lawfully obtain digital copies of books. Under negotiated terms, publishers provide their ebooks to library aggregators, who, in turn, work with libraries and their patrons to host a platform that permits and tracks the lending of ebooks. In this way, library patrons can log onto their library accounts from their homes or any other location, download ebooks onto their computers, smartphones, or e-reading devices for a limited time period under stated terms of use and read them without physically visiting their local library. Each library focuses its services on its community, making ebook (and other) acquisition decisions based on the specific needs of local patrons.

48. As competitors in the marketplace, Plaintiffs have each worked with libraries and library aggregators to develop innovative models capable of sustaining a functioning market for their respective ebooks. These pricing and agreed terms of use—which vary from publisher to publisher and can vary by market segment—continue to evolve based on the feedback from libraries and other considerations. There is a vibrant market for selling and licensing ebooks to libraries to provide their patrons with lawful copies of ebooks. But that market cannot be sustained if, rather than patronize their local library, individuals can freely download unauthorized scanned copies of Plaintiffs’ books from IA’s Website.

iv. Plaintiffs and Libraries Reacted Rapidly to Ensure Library Patrons Have Access to Books During the COVID-19 Lockdown

49. The advent of the COVID-19 pandemic has shown that online commerce is more important than ever to publishers, authors, bookstores, and libraries. During this time period, many customers have been well-served by existing digital business models, both commercial and noncommercial. Additionally, publishers and libraries have engaged in numerous emergency-related initiatives to ensure that readers retain access to books while nationwide stay-at-home orders remained in effect. They have ranged from providing free ebook copies to library patrons during the shutdown in specific instances, to donating hundreds of thousands of print books, to working closely with schools, colleges, and teachers to ensure access to and availability of books necessary for online learning.

50. The Publishers have used their exclusive rights to strike an appropriate balance between the interests of the relevant parties, particularly authors and readers. Moreover, the measures taken by the Publishers in response to the COVID-19 pandemic also take into account the needs of booksellers, particularly independent bookstores, most of which are small businesses that have been crippled by the shutdowns. Copying and giving books away to

everybody for free—as IA has done with Plaintiffs’ and others’ works—deprives booksellers of the sales that they need to stay afloat and authors of royalties they otherwise count on as compensation for their work.

51. Libraries have been equally proactive in meeting the challenges of the COVID-19 pandemic. For example, despite the closure of their physical locations, many libraries have allowed new patrons to sign up for library cards online. Patrons are still able to lawfully download hundreds of thousands of ebooks and online materials through the digital lending systems that libraries and publishers have developed. IA’s Website steers readers away from the digital platforms that local public libraries continue to operate.

B. Internet Archive Unlawfully Disrupts the Book Publishing Ecosystem by Infringing Copyright on an Industrial Scale

52. Despite its efforts to cast itself as the hero in this story, IA’s business model for the Website—which is essentially to freely disseminate scanned copies of every physical book it can lay its hands on—is parasitic and illegal. IA exploits the invaluable work that authors and publishers do without investing in any of the effort or paying any of the costs associated with the creation and publication of the books. What IA does is copyright infringement, plain and simple, and it must be stopped.

i. Internet Archive

53. Brewster Kahle founded Internet Archive in 1996. Internet Archive provides a number of services not at issue in this action, including its Wayback Machine and digitization of public domain materials. At issue here is IA’s scanning of in-copyright books and distribution of digital copies via its Website. Notably, it operates with a surprisingly unsubtle penchant for money-making behind its non-profit 501(c)(3) tax code designation.

54. IA reported more than \$150 million of revenue in the last ten years, according to publicly available tax filings. As per its 2017 tax filings, it employed 150 employees. IA's headquarters are located in an exclusive area of San Francisco. Kahle expanded the IA empire in 2019 by purchasing through Better World Libraries, a shell company he controls, the for-profit Better World Books, an online retailer that predominantly sells used books.

55. The bulk of IA's revenue is derived from contributions from large donors, including tens of millions of dollars from the Kahle/Austin Foundation. The Kahle/Austin Foundation is an entity established by Brewster Kahle and his wife to give money to IA and other favored projects. In 2018, the Kahle/Austin Foundation reported assets of \$104,483,456. IA also has reported sizeable donations from other large foundations, some of which are based in New York.

56. IA has an interlocking web of contributions and commercial services that support its Website. In addition to receiving large-dollar donations, IA has made tens of millions of dollars from selling commercial services. One of the services it offers is industrial-scale book scanning and digitization, which has generated more than \$25 million in revenue since 2011. IA provides this service to customers nationwide, and as its marketing materials tout, employs a Regional Digitization Manager for customers in the "NJ/NY/PA" area. Upon information and belief, this employee currently resides in New York City. This same employee helped to set up and manage IA's first digitization center, which was housed for an extended period of time in this District.

ii. The Open Library

57. IA started the "Open Library" in or around 2006 with the goal of providing "one web page for every book published." An online catalog of all the books in the world was not a novel concept. Since 1998, the WorldCat database run by the Online Computer Library Center

has provided an exhaustive online catalogue of library books, which currently contains over 450 million bibliographic references.

58. But IA was not content with a resource tool designed to help users find books. Taking a step further, IA created the Website with the goal of providing free downloads of entire copies of every book ever published. Over the course of several years, IA has gradually built up a program designed to obtain vast quantities of print books in bulk, scan them on an industrial scale, and distribute digital copies through the Website without any license from the copyright owner.

59. IA engages in this massive, industrialized scanning of print books to create digital files in large part to evade the Publishers' commercial terms for ebooks and because DRM precludes the duplication of Plaintiffs' ebooks. In other words, IA employs this end-run as a means to avoid both the Publishers' ebook use restrictions and the dictates of the DMCA by creating its own bootleg electronic versions of the books through scanning. But that end-run is unlawful and equally harmful to the Publishers and their authors, who receive no compensation for IA's reproduction and distribution of their works in digital form and whose paid offerings cannot readily compete with IA's free but unlawful versions.

60. The basic purpose of IA's massive book digitization project is the same as Plaintiffs' basic purpose in publishing books, which is to distribute reading material. This is in stark contrast to other large-scale book digitization projects, like the ones carried out by Google and the HathiTrust, which created digital indexes of the contents of books that could be used for the purpose of online search, but which never made the contents of entire books freely available to the general public. Here, IA allows every person in the world to instantly download complete

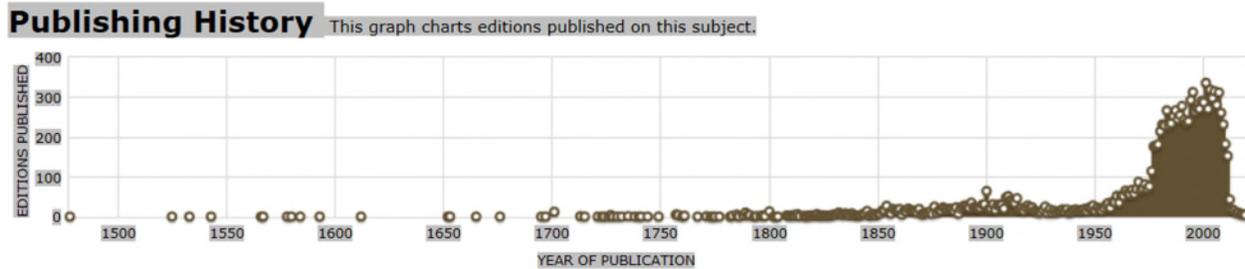
books or download them—without ever receiving the necessary permission from the copyright owner.

61. In a 2017 interview, IA’s founder stated, “We’re trying to . . . make all the published works of humankind available to people, permanently. If you’re curious enough to want to have access, we can make it available [sic] to all the books, music, video, web page, software, lectures, available to anybody wanting to have access.” Mary Kay Magistad, *Where to find what’s disappeared online, and a whole lot more: Internet Archive*, PRI.org (Feb. 23, 2017, 8:00 p.m.), <https://www.pri.org/stories/2017-02-23/where-find-whats-disappeared-online-and-whole-lot-more-internet-archive>.

62. In recent years, IA has started to ramp up its efforts. The annual income reported by IA has nearly doubled since 2013 and the rapid accumulation of funds has been accompanied by public statements promising aggressive expansion. In 2019, for instance, the Director of Open Libraries told *Library Journal* that he aimed to increase the acquisition of books in bulk and speed up the rate of book digitization to 500,000 books per year. The goal of this accelerated activity is to put “more than four million books online for the public,” which would match the number of books in the collection of a large metropolitan library system. Matt Enis, *Internet Archive Expands Partnerships for Open Libraries Project*, *Library Journal* (May 2, 2019), <https://www.libraryjournal.com/?detailStory=internet-archive-expands-partnerships-for-open-library-project>.

63. As the number of scanned books on the Website increases exponentially, the misrepresentations IA employs to justify its infringing activities are exposed. For example, IA frequently characterizes the books on the Website as twentieth-century works that do not have active sales or available ebooks, as if infringement is allowed if a work is not readily available

(which it is not). But this “old books” defense is a fantasy. As is evident from IA’s own figures and the Works in suit, many of the books that IA provides in the most popular categories, such as biographies, were published in the years leading up to 2000 and in the twenty-first century, with relatively few books from earlier in the twentieth century. This is evident from the Open Library’s own graph illustrating the year of publication for its biographies:



64. Indeed, the Works in suit include books first published in 2019, including *The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution* by Gregory Zuckerman, with a publication date of November 2019.

65. In short, IA directly harms the Plaintiffs’ print *and* ebook markets in all market segments by providing competing substitutes for numerous original works currently available in their catalogs.

66. IA knows that it cannot proceed with the mass digitization and distribution reflected by the Website without an agreement with the Publishers on mutually agreed terms. The Publishers and/or their trade association AAP have put IA on direct notice that Open Library’s unauthorized reproduction and distribution of copyrighted works is infringing. For example, in 2018, HarperCollins put IA on express written notice that its actions in connection with Open Library with respect to HarperCollins’ titles were infringing and not justified under its various manufactured defenses. Despite actual notice that its actions were illegal and without any basis in law, IA has willfully persisted with its infringing activities.

iii. The User Experience of the Website

67. From the perspective of the Website user, IA offers an ostensibly attractive proposition: free copies of about 1.3 million books. Anyone with an Internet connection and an email address can sign up for an account to use the Website in a matter of minutes. The process requires no verification of the user's identity, and there is nothing to prevent users from creating multiple accounts with dummy email addresses in order to circumvent limits, assuming any are in place at all.

68. Once on the Website, a user is taken to a landing page that suggests titles under categories such as "Books We Love," "Romance," "Science," "Kids" and "Thrillers." Users can also search for books according to title, author, subject, ISBN, and other categories.

69. Clicking on a book title takes the user to the title's webpage on the Website, which contains a picture of the cover, a description of the book and bibliographic information, such as the publication date and the WorldCat catalog number. As shown in the screenshot on the following page, which is a typical webpage on the Website, the user is also presented with links through which he or she can buy a copy of the book.

The screenshot shows the Open Library website interface. At the top, there's a navigation bar with 'OPEN LIBRARY' logo, 'Browse', 'My Books', 'More', a search bar, and 'Log In'/'Sign Up' buttons. Below this is a blue banner for 'K-12 Student Library'. The main content area features the book 'Blink' by Malcolm Gladwell. On the left is the book cover. To the right of the cover, the title 'Blink' is displayed in a large font, followed by the subtitle 'the power of thinking without thinking' and the author 'Malcolm Gladwell'. Below the author name, there are star ratings (3.33), a 'Want to read' count (122), and a 'Currently reading' count (5). There are also links for 'Borrow', 'Preview', and 'Download for print-disabled'. A 'Buy this book' section lists 'Better World Books' as the first option with a price of \$3.98 (used) including shipping, followed by 'Amazon' at \$5.68 and 'Bookshop.org'. There are also social media sharing options for Facebook, Twitter, Pinterest, and an Embed button.

70. The first—and most prominent—“Buy this book” link is to Better World Books, the online, largely used bookseller owned by a shell company controlled by IA founder, Brewster Kahle. Notably, IA does not provide links to the book publisher’s or author’s website.

71. For each book on the Website, IA gives users one of two options, which it calls “read” or “borrow.” The books in the “read” category are books that IA presumably has concluded are in the public domain. For books in the “read” category, the Website has no limit on how many users can download them, either at any point in time or in the aggregate. Users who choose the “read” option can download to their computers, e-readers, or mobile devices a scanned copy of the entire book in a variety of digital formats that lack any DRM protection,

including as a .pdf file. Once a user downloads a book as a .pdf, he or she is free to make unlimited copies of the book and can distribute it to whomever they wish at no cost.

Alternatively, users can read these books through the IA's "Book Reader" interface described below.

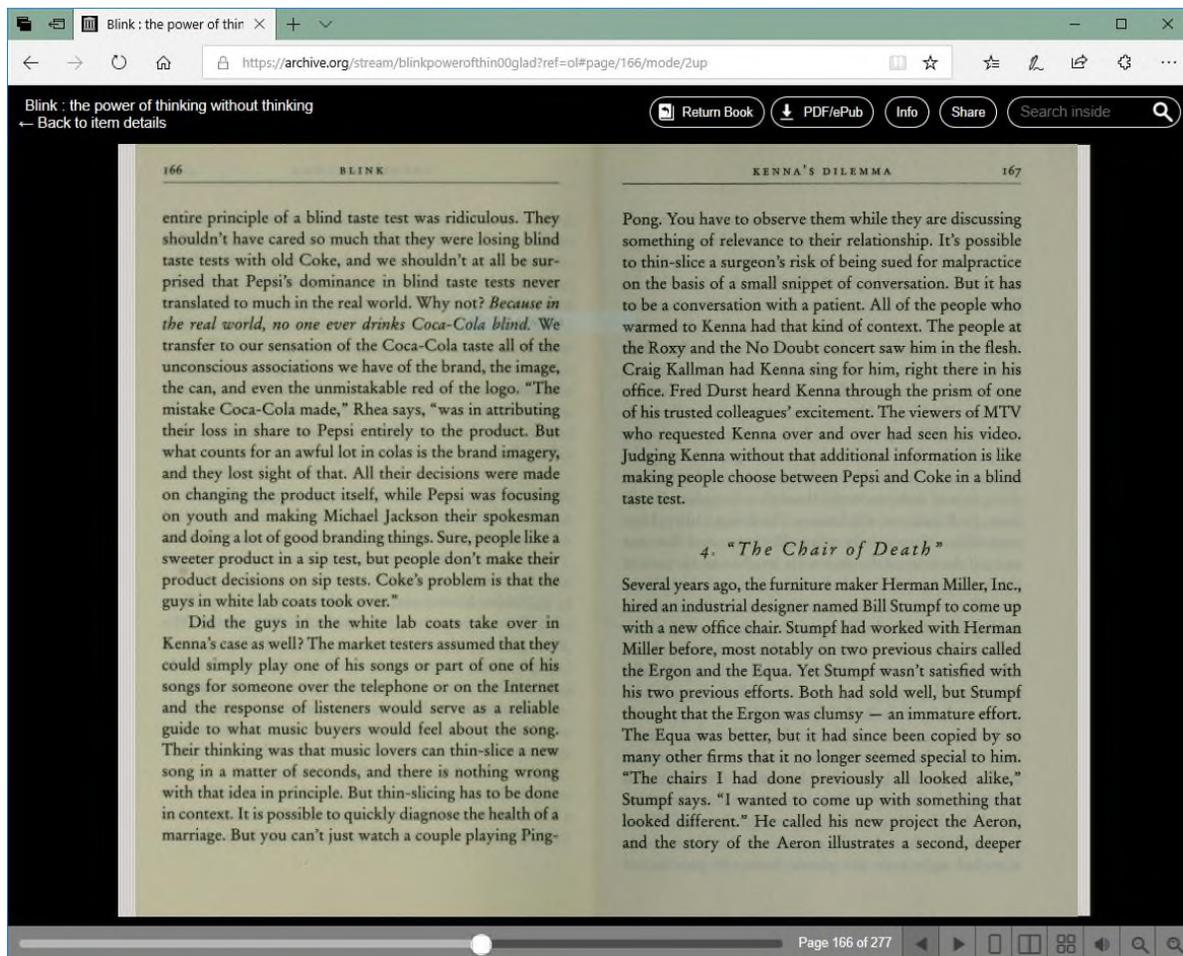
72. Although the books that IA classifies in the "read" category are presumably supposed to be in the public domain, sometimes the books are in fact protected by copyright. For instance, at one point, any Internet-connected individual could download copies of the modern classic *To Kill a Mockingbird*, which is still under copyright, without any restrictions. In other cases, copyrighted books have been distributed without DRM restrictions because, upon information and belief, IA erroneously decided that they were public domain works based on elementary misunderstandings of copyright law.

73. Under the second option, IA permits users to "borrow" titles from the Website that it recognizes are not in the public domain through its so-called "controlled digital lending" protocols. "Borrow" is a euphemism for an illegal reproduction and distribution. As described below, "controlled digital lending" is a manufactured legal paradigm, conceived by IA, to cast aside well-established copyright jurisprudence.

74. Under IA's "controlled digital lending," a Website user can download a certain number of books at a given time, which IA has currently set at ten. Once IA has distributed the maximum number of books the user is allowed, he or she must check a book back in before taking another one out, although there is nothing to prevent a user from circumventing this limit by setting up multiple accounts. Each such "loan" lasts for a two-week period. When the period ends, the book is supposedly "checked in" by the Website.

75. Until the recent pandemic and the advent of IA’s so-called “National Emergency Library,” IA claimed that it also enforced an “owned to loaned” ratio that restricted the number of users who could borrow a copyrighted book at once. In theory, this means that the number of scanned copies of a title downloaded from the Website at any one time cannot exceed the number of print copies of that title owned by IA or its partner libraries. If that number is exceeded, then the user is ostensibly put on a “waitlist.”

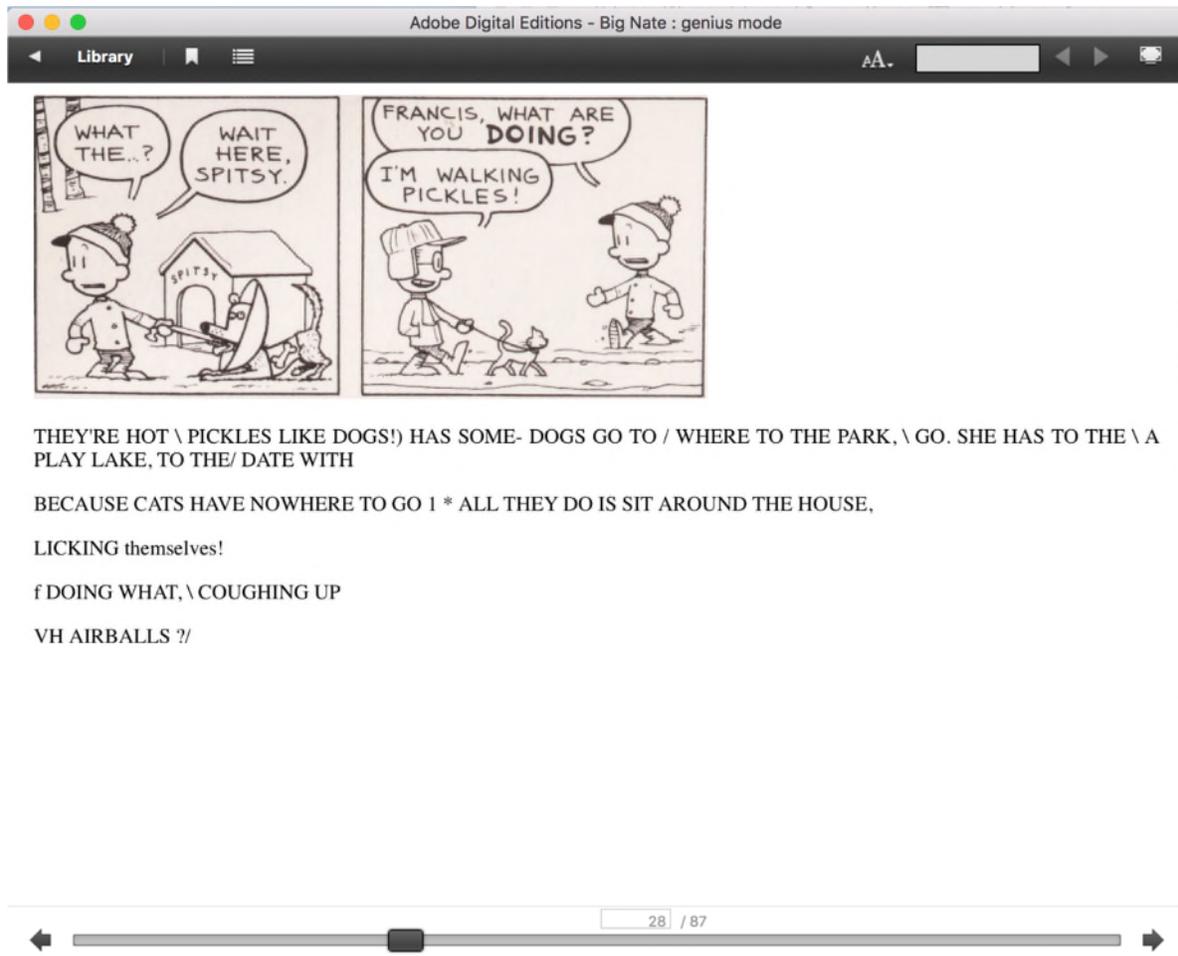
76. Users that have successfully checked out a book can read it immediately on the Website’s Book Reader platform. IA publicly displays a copy of the work on that platform. This interface operates in the user’s web browser and provides a digital replica of a hard copy book, as shown in the screenshot below:



Users can turn the pages by clicking on them or using buttons below the page. The Book Reader also allows users to click an “audio” button, at which time IA’s software mechanically reads the book aloud.

77. IA also enables users to download and view copies of books using a software program called Adobe Digital Editions (“ADE”). The ADE files are purportedly protected by DRM technology that restricts certain copying and only allows users to access the file during the fourteen-day loan period, at which time it locks them out.

78. The Website allows users to “borrow” books in two ADE formats: “Encrypted Adobe PDF” (which is billed as containing “High Quality Page Images” of the entire hard copy book) and “Encrypted Adobe ePub.” The Encrypted Adobe PDF is not the kind of specially-formatted ebook that a user would purchase as an authorized ebook. Rather, it is a digital file consisting of a scan (*i.e.*, photograph) of the pages of a physical book. IA generates the Encrypted Adobe ePub file from scanned books using optical character recognition technology. The Website distributes the Encrypted Adobe ePub as a “smaller file,” but acknowledges it “may contain errors.” This is an understatement. The ePub files tend to be rife with transcription errors and are sometimes entirely illegible, as can be seen from the example on the following page.



79. Users can read the books they download as ADE files on devices such as smartphones, tablets, and e-readers like Kindle, Nook or iPad. The Book Reader platform has been designed to work on mobile devices as well. IA has boasted on its blog that Website users can borrow books and read them on their devices just as easily as they can read ebooks legitimately purchased or borrowed from licensed services. While the quality of the digital format scans that IA provides are inferior to the quality of Publishers' ebooks, these bootleg versions act as a substitute for the authorized versions, since readers select titles for their content. No one reads a James Patterson thriller after downloading a scan of the book from Open Library and then declares, "I liked it so much, I am going to read an authorized ebook again on my Kindle for a different experience."

iv. The Open Library Is Not a Library, It Is an Unlicensed Aggregator and Pirate Site

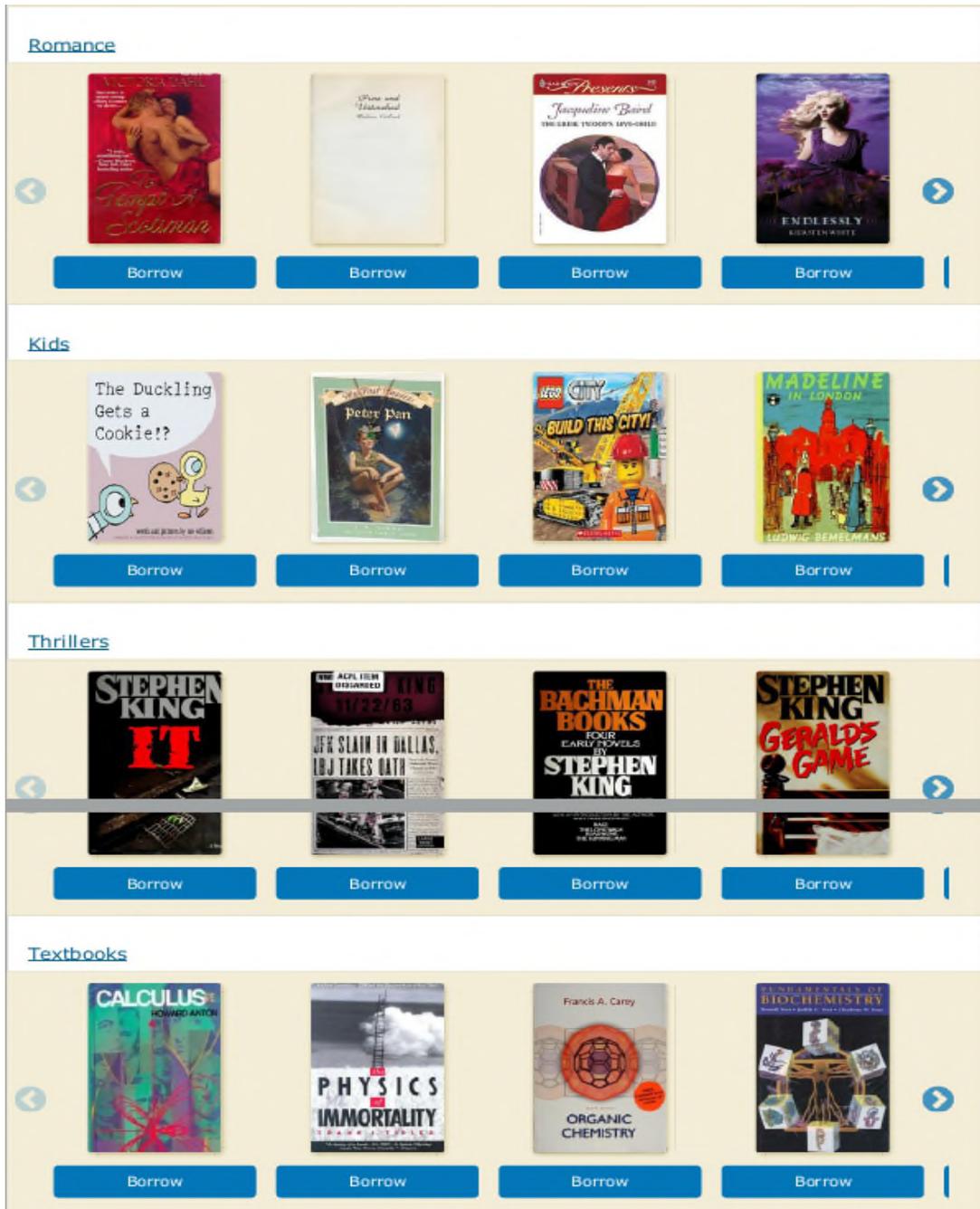
80. Defendant bills its Website as “an accredited California State Library run by the non-profit Internet Archive,” but this branding fraudulently misleads on several levels.

81. The Open Library is not an “accredited library” by any commonly understood definition of those words. The reference to accreditation refers, upon information and belief, to the fact that the State of California provided IA with federal funding through the Library Services and Technology Act (“LSTA”) in or around 2011. LSTA awards are available to a wide array of organizations, not just actual libraries, including for example the Los Angeles Philharmonic Association. In short, the fact that Internet Archive was able to meet the broad criteria for LSTA funds does not make it an “accredited library,” and it most certainly does not make it the trusted equivalent of a public or academic library.

82. More fundamentally, the Website lacks the characteristics shared by actual libraries. Instead, the service it provides more closely resembles that provided by aggregators like OverDrive, that Plaintiffs routinely engage to distribute ebooks, with defined terms of use and at an agreed-to fee—except IA operates without authorization or remuneration. For instance, whereas libraries serve local and academic communities, the Website distributes copies indiscriminately to everybody on the planet, or at least everybody with a connection to the Internet.

83. IA tries to wrap itself in the flag of an “educational” enterprise, but it is not. First, it is not an educational institution or even an affiliated academic library serving a defined university community. It is devoted to education no more specifically than any platform that transmits, streams, or otherwise delivers content on the global Internet. Moreover, the notion that the Website specializes in educational material is pure fiction. A quick glance at the

Website’s landing page shows that it lists thrillers and romance novels more prominently than textbooks or other works that may be used in schools. Further, upon information and belief, many of the non-fiction books on the Website are read for personal entertainment or edification, not classroom use:

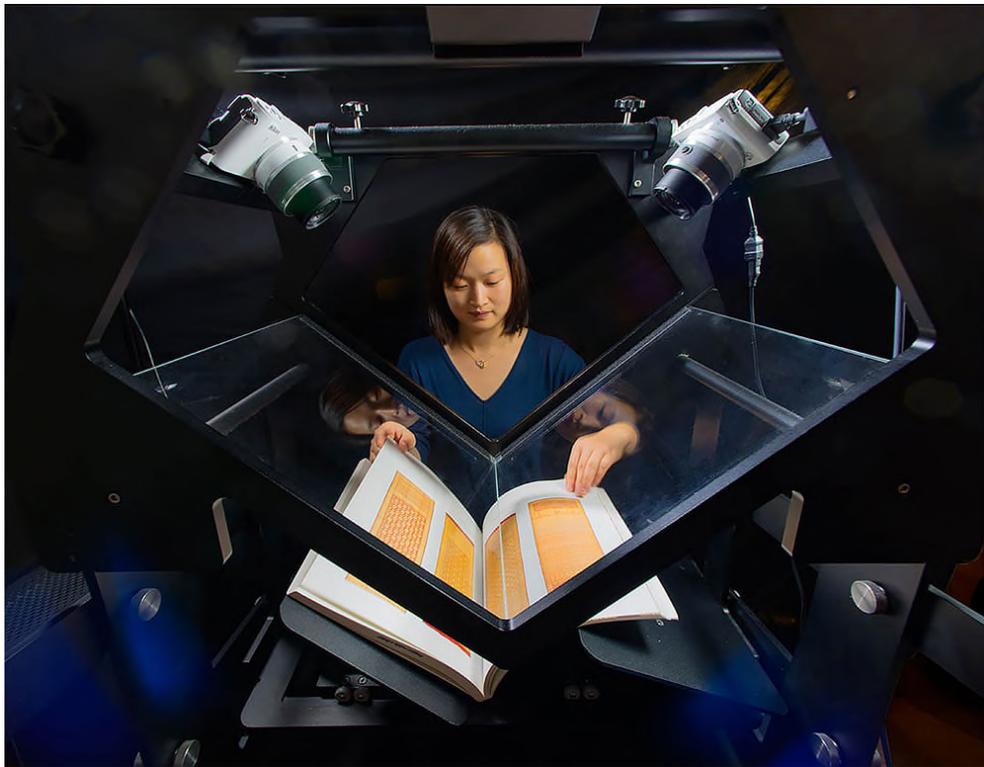


84. Branding itself as a “library” does not imbue Defendant with the right to engage in behavior that would otherwise be considered wholesale theft.

v. IA’s Industrial Book-Scanning Machine and Global Scanning Operations

85. IA has devised a commercial full-service loop that funds its quest to distribute digital copies of every book, and at the same time, provides millions of copies of the very works it needs to stock its Website.

86. To this end, IA has implemented an industrial process that enables it to scan the many millions of books it has acquired and continues to amass. The “Scribe system” developed by IA uses a sophisticated scanner operated by manual labor. *See* Internet Archive, Internet Archive Digitization Services – Partner Documents, <https://archive.org/details/partnerdocs> (last accessed May 31, 2020). IA invented the Scribe scanner for the purpose of scanning books in bulk:





87. Upon information and belief, a single Scribe scanner can digitize a *300-page* book in five minutes. According to its promotional materials, IA operates seven “regional digitization centers” and eleven “satellite locations” in the United States, Canada, and the United Kingdom, with multiple Scribe scanners at each location. Upon information and belief, IA has employed more than 50 “book scan center staff” to work within facilities housing Scribe scanners, who are responsible for digitizing the books to add to the Website. Upon information and belief, IA has also engaged low-wage contractors in other countries, such as the Philippines and China, to scan bulk quantities of books that are shipped there from the United States. Upon information and belief, IA also has recruited and used unpaid volunteers to scan copyrighted books for the Website. IA can scan 3,000 books in a single day when operating at full capacity, according to its Archivist and Software Creator, Jason Scott (believed to be a New York resident).

88. At one point, IA operated a digitization center in this District and, although it has been discontinued, the Website currently contains many copyrighted books that were scanned in New York, including some of Plaintiffs' Works in suit.

89. IA makes multiple reproductions of Plaintiffs' books at various stages. As IA photographs each page of a book using the Scribe scanner, the scanner creates a digital copy of those pages and then combines them with copies of all the other pages to make a single digital copy of the entire book. IA then copies each digital book file onto its main computer servers located in San Francisco, together with metadata about the book and the scan or file. Multiple copies of the scanned book pages are made in the course of this process as IA converts the file into different formats (including, *inter alia*, .pdf, ePub, OCR text files and audio formats) and then reproduces all of those files onto its servers.

vi. Internet Archive Shifts the Costs of Its Illegal Book Digitization Project Onto Publishers, Libraries, Open Library Users, and Other Third Parties

90. IA's business model for Open Library is predicated upon infringement, and it willfully inverts the carefully balanced book publishing ecosystem. Whereas publishers, authors and libraries actually invest the time and money that it costs to publish and distribute a book, IA refuses to pay for any of these costs and has even devised a parasitic system to shift the costs of its large-scale infringement onto other people.

91. IA attempts to present the Website as an altruistic non-commercial enterprise. But despite its technical not-for-profit status, IA has set itself up to generate huge amounts of revenue from the very people and entities it is supposed to be helping in order to fuel its copyright infringement project, as the following examples illustrate. IA should be seen for what it actually is: a commercial actor.

a. Internet Archive is Paid Millions of Dollars to Infringe Copyrighted Books and Put Them on the Open Library

92. IA operates a major commercial business performing scanning and hosting services. This provides a critical revenue stream that IA uses to fund and facilitate its piracy scheme. IA first induces various libraries to pay to have books in their collections scanned at a set price, and, at the same time, copies of all the works scanned are provided to the Open Library. While most public libraries have agreed to scan only public domain materials, on information and belief, IA also has earned substantial fees scanning and uploading some library collections of copyrighted books.

93. In a recent blog post, Brewster Kahle admits that “[l]ibraries paying for our scanning services is a major source of earned income for the Internet Archive.” *See* Brewster Kahle, *Internet Archive Staff and Covid-19: Work-at-Home for Most, Full-Pay Furlough and Medical for Scanners* (Mar. 25, 2020), <https://blog.archive.org/author/brewster/>. Evidence from IA’s publicly available tax records confirm that it has made over \$25 million from book scanning since 2011. Of course, IA also profits from adding copies of the books it scans to the Website, using funds from donations, foundations, and elsewhere.

94. Plaintiffs obviously have no objection to the digitization of public domain titles. Nor do they object to anyone granting IA a license to put books or journals that they own the rights to on the Website. But IA has designed its bulk book scanning service for libraries to be a pipeline that feeds the Website with extensive copyrighted material in addition to the permitted copying of any public domain or licensed material, all at no cost to IA.

b. Better World Books Feeds the Open Library with Copyrighted Books

95. The acquisition of Better World Books by a shell company controlled by Brewster Kahle in 2019 has provided IA with yet another way to acquire books for the Website without

actually paying for them. As part of its business model, Better World Books acquires mass quantities of used print books, including what it refers to as a “Library Discards & Donations” program that it claims works with librarians to “reuse or recycle” surplus library books. Better World Books funnels books to IA, which scans and uploads them to the Website. Moreover, the complex series of circular relationships between IA and Better World Books further contributes to the commercial nature of IA.

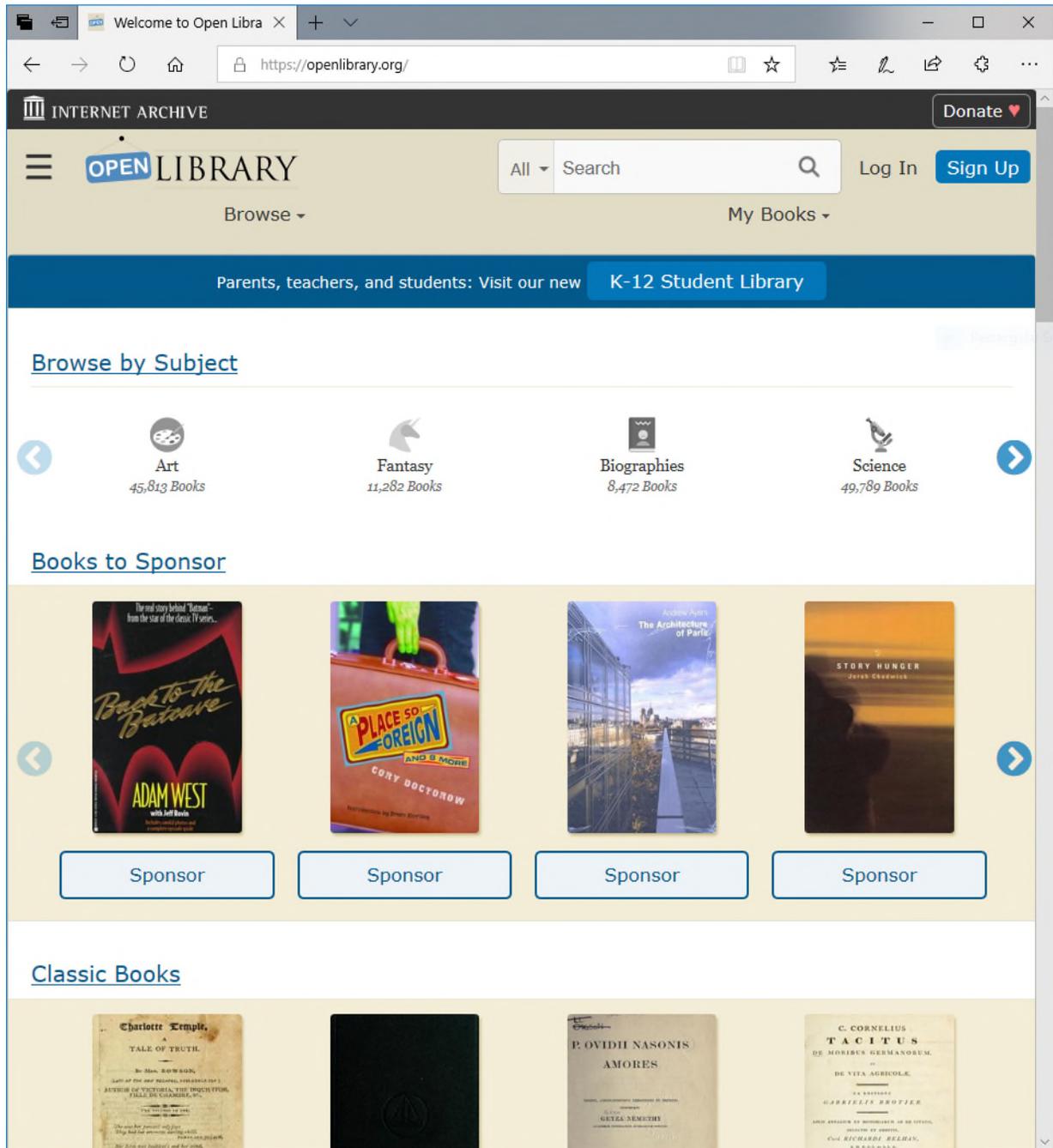
96. This arrangement creates a perfect closed loop system that benefits IA in several ways. As IA stated in the press release accompanying the acquisition, the new relationship between Better World Books and IA “will allow Better World Books to provide a steady stream of books to be digitized by the Internet Archive, thereby growing its digital holdings to millions of books.”

97. Upon information and belief, the system works as follows: first, Better World Books acquires millions of used print edition books (provenance unknown), which it provides to IA to be scanned and added to the Website for widespread downloading and distribution. Second, the webpage for every digital format book on the Website—including the pages for the books provided by Better World Books—includes a prominent link inviting the user to buy a print copy of the book from Better World Books. Upon information and belief, this sale not only benefits Brewster Kahle’s shell corporation, but the traffic that IA drives to Better World Books results in more book sales, which enables IA to acquire more books to scan and upload to the Website. And the cycle repeats. Better World Books also encourages its customers to donate funds to IA by topping off their purchases to a higher dollar amount.

c. Users “Sponsor” Books that Internet Archive Wants to Infringe

98. IA also earns revenue from users in exchange for special access to individual book titles through its “sponsorship” program. In some ways, this is little different from

“selling” limited access to digital books to these sponsors. In brief, IA gives users the option to “sponsor” books that it does not have and wants to add to its collection. The sponsorship program is so central to IA that the main landing page for the Website features at the very top the selection of books it would like users to sponsor:



99. Users who click on the sponsor option are taken to a webpage that informs them that a “tax deductible donation can add this book to Internet Archive’s lending library, forever.” The book page informs the user of the cost of the “donation” required to purchase the book.

The screenshot shows a web browser window displaying the Internet Archive donation page for the book "Back to the Batcave" by Adam West. The page features a navigation bar with links for ABOUT, CONTACT, BLOG, PROJECTS, HELP, DONATE, JOBS, VOLUNTEER, and PEOPLE. The main content area includes a headline: "Your \$93.12 tax deductible donation can add this book to Internet Archive’s lending library, forever." Below this is a book cover for "Back to the Batcave" and a list of sponsorship benefits: "Buying & shipping Back to the Batcave", "Beautiful high-fidelity digitization", "A personalized sponsorship message", and "Notification by email to claim first borrow*". A note states "* Most books available within 1 month". There is a section for "Your sponsorship preferences" with a message: "If you want to personalize a sponsorship message, and be first to borrow this book, you must be logged in." A summary table shows "Sponsorship total" as "\$93.12 USD". On the right, there is a "Sponsor this book" button and a "Pay with a credit card" button. The footer contains tax information: "The Internet Archive is a US 501(c)(3) non-profit organization. Our tax ID number is 94-3242767. Your gift is tax deductible to the full extent provided by the law."

100. In this case, users were given the option to purchase for IA a copy of *Back to the Batcave* by the late actor Adam West, who played Batman in the cult-classic 1960’s television show, for a hefty donation of \$93.12. As an inducement to sponsor the book, users are given the right to be the first to borrow the book when it becomes available and to add a “personalized sponsorship message.” IA uses donations it receives for sponsored books to obtain a physical

copy of the book, scan it, and upload the digital copy to the Website for download. It is not clear how IA uses any money left over.

101. IA's book sponsorship scheme is breathtakingly brazen. In essence, IA tells its users which copyrighted books it wants to infringe. Then it asks users to pay a "donation" far in excess of the list price of the book for IA to go out and buy a print copy from an undisclosed source, possibly Better World Books. Finally, IA uses the user's money to scan the book and put it online, where anyone can get a copy for free—completing the copyright infringement process without spending a single dollar of its own money. Tax-deductible donations are not meant to foster piracy.

102. The sponsorship option is also designed to drive an accelerating cycle of infringement. By encouraging users to sponsor the books they most want, IA incentivizes its users to select popular books. Once IA obtains the print book, scans it, and uploads it to the Website, that infringing file will attract new users. The new users will, in turn, sponsor even more popular books, which will attract even more users and so on, *ad infinitum*.

d. Internet Archive Absorbs Entire Libraries for the Website

103. Yet another way IA acquires books for free is to solicit and accept bulk donations of books from struggling or defunct libraries. In a May 2019 article, Open Libraries Director Chris Freeland described IA's "acquisition program," stating that it "involves collecting donations of *in-copyright* materials from libraries and booksellers, digitizing this content in [IA's] scanning center in Cebu, the Philippines, and then storing the physical copies in archive facilities in Richmond, CA" (emphasis added).

104. Ironically, the many thousands of hard copy books IA obtains from defunct colleges or libraries will likely end up in "archive facilities in Richmond, CA," which consist of large shipping containers owned by IA. Once locked away, upon information and belief, IA will

make no effort to make the print books available to be read, like books in actual library collections. Instead, the print copies primarily exist to rationalize, or provide the predicate for, IA's argument that there is a one-to-one correlation between print copies legitimately owned and their illegitimate ebook scanned copies.

105. Through its acquisition program—and similarly aggressive book scanning, Better World Books and “sponsorship” schemes—IA seeks to achieve its next benchmark of putting 4 million copyright-protected books on the Website.

106. Remarkably, every one of the systems for acquiring books described above is designed so that IA avoids paying authors or publishers anything, while simultaneously ensuring that IA rakes in money from its infringing services and, at the same time, obtains free books for the Website.

vii. Controlled Digital Lending

107. IA relies on the contrived theory of “controlled digital lending” described above as its central defense to charges of copyright infringement. But CDL is an invented paradigm that is well outside copyright law, appears to have sprung up in response to the objections of copyright owners to IA's infringing activities and, in any event, does not excuse IA's massive infringement.

108. IA is a leader and organizer of a larger proselytizing movement of academics and activists seeking to find a way to justify a “scan first, figure out the details later” approach to the mass digitization of copyrighted books. In 2018-19, IA sponsored the drafting of a “White Paper on Controlled Digital Lending of Library Books” by law professors David R. Hansen and Kyle K. Courtney. IA and Brewster Kahle broadly promote CDL, speaking widely and engaging in a broad public relations campaign, inducing others, including libraries, to join their cause and

cynically using them to reflect their glow of legitimacy onto the Website. But IA's "controlled digital lending" finds no actual support in the law.

109. As a preliminary matter, Plaintiffs' challenge whether IA maintains the detailed records or practical control necessary to sustain the so-called "owned to loaned ratio" that is the cornerstone of CDL, *i.e.*, the notion that the number of electronic copies of a book for download from the Website at a given time never exceeds the number of physical copies of the book owned by IA or one of its partner libraries. With respect to the Website's titles for which the corresponding print books are allegedly stored at partner libraries, it defies reason that the partner libraries will have the wherewithal to faithfully and consistently remove a book from circulation each time it is borrowed on the Website, and put it back on the shelf when the Website version is checked back in. As for the truckloads of books warehoused by IA, IA archivist Jason Scott admitted in a recent tweet that for the millions of physical copies acquired by IA, "[o]nly one or two unique copies are kept" and "stored at the physical archive." "[D]uplicates (which have increased over time) are donated to various charities and non-profits."

110. But even if IA were scrupulously following its own invented theory of "controlled digital lending," the theory has no legal justification. First, IA and its supporters relied heavily on the "first sale doctrine" codified at 17 U.S.C. §109, which entitles the lawful "owner of a particular ... lawfully made" copy of a copyrighted work, like a book, "to sell or otherwise dispose of the possession of that copy or phonorecord." The central thesis of the White Paper sponsored by IA was that the broader principles reflected in this doctrine should be imported into the fair use doctrine to protect IA's actions. But as enacted by Congress, the first sale doctrine is carefully confined as a limitation on only the distribution right. It permits the owner of a copy to *distribute* the particular copy that has been lawfully acquired—for example, as in the secondary

sale of a hardcover book or a painting—but it provides no exemption from the copyright holder’s exclusive right to *reproduce* a work. The lynchpin of IA’s whole operation is that it scans a print book to create a digital file—a classic unauthorized *reproduction* of a work that puts the application of Section 109 clearly out of reach.

111. Faced with these inconvenient facts, IA has also advanced the illogical premise that its massive illegal copying and distribution is “format shifting” protected by the fair use doctrine. But the rudimentary use of a scanner to “format shift” print books into digital works for distribution to the public does not constitute either a permitted personal use or a transformative use. The activity squarely intrudes upon the exclusive rights under 17 U.S.C. § 106, and no exception of any kind applies.

112. At bottom, CDL is based on the false premise that a print book and a digital book share the same qualities. But, as outlined above, they are fundamentally different mediums, and they exist as distinct economic markets. As the Copyright Office phrased it in a key report, “Time, space, effort and cost no longer act as barriers to the movement of [digital] copies, since digital copies can be transmitted nearly instantaneously anywhere in the world with minimal effort and negligible cost.” Digital Millennium Copyright Act (DMCA) Section 104 Report: Before the Subcommittee on Court, the Internet and Intellectual Prop., 107th Cong. (Aug. 2001) at p. 82, available at <https://www.copyright.gov/reports/studies/dmca/sec-104-report-vol-1.pdf>. In stark contrast, a range of physical obstacles impact the distribution of books as material objects—from the need to transport them to each reader, to the need for library patrons to travel to libraries to take them out and return them. Further, print books deteriorate over time, unlike digital files. Copyright law vests the copyright owner with the right to develop each market based on terms appropriate to the medium, as well as the right to extract the full value of

publication in digital form. These economic rights would be severely undermined if IA can circumvent the rightsholder entirely and copy millions of print books into digital copies to be widely distributed, even if it maintains an “owned to loaned ratio.”

113. IA’s mass digitization of books is potentially even more pernicious than ordinary online piracy. First, its use of “library” branding deceives some users into thinking the Website is a legitimate site. Moreover, IA not only makes millions of dollars as outlined above in ways related to its systematic infringement but also uses those profits to fuel its mission to fill the digital Library of Alexandria with every book ever written. Allowing IA to operate the “Open Library” for in-copyright works will create an indelible impression for libraries and readers that digital format books should be free—and can be free just as soon as IA scans a print copy. This, in turn, will impair the sales and licensing of books and ebooks that actually make it possible to pay the high cost of writing and publishing the quality books that IA purports to value so much, yet pays little or nothing to sustain.

viii. The National Emergency Library

114. On March 24, 2020, IA doubled down on its infringement, announcing that in light of the COVID-19 pandemic, it would suspend the “owned to loaned” ratio aspect of CDL, while retaining other aspects of it. In other words, Defendant altered how the Website operates so that it could distribute as many copies of any book it wanted, no matter the number of print copies on hand purportedly because the pandemic required it. This so-called “National Emergency Library” was tantamount to asserting an emergency copyright act unilaterally and by private action.

115. IA admits that it did not bother to “engage with the creator community and the ecosystem in which their works are made and published” before launching this revised approach for the Website. When IA was roundly criticized by publishers, authors, individual librarians,

and the bookselling community for disregarding their interests and legal rights, IA responded that it had merely “moved in ‘Internet Time’ and the speed and swiftness of our solution [to the COVID-19 lockdown] surprised some and caught others off guard.”

116. In a feigned act of magnanimity, IA assured authors that it would abide by a notice and takedown system. But this turns copyright law on its head. Copyright owners have the power to decide in advance how their exclusive rights will be exercised. Copyright is not an “opt-out” system whereby infringers can distribute copyrighted works for free, with impunity, until they are told to stop. The Copyright Act does not and never has put the burden on authors and publishers to police the unlawful actions of direct infringers, which in the case of the IA not only copies, uploads, and distributes infringing files, but asserts conditions and procedures for agreeing to stop the infringement.

117. Defendant does not qualify as an intermediary entitled to certain protections under the Digital Millennium Copyright Act. The DMCA allows the operators of online services that host content *posted by users* and meet certain conditions to avoid copyright liability by responding expeditiously to a takedown request from the rightsholder. The DMCA notice and takedown system is inapplicable to this case because IA, not third-party users, uploads the infringing files to the Website. And to add insult to injury, both before and since the “National Emergency Library,” when faced with the frustrated demands of authors or publishers, IA has often failed to stand by its proclamation to honor takedown requests, or opt-outs, thereby further exposing the overall lack of integrity that surrounds this process.

118. The Website as operated under the “National Emergency Library” parameters does not qualify as a fair use. Among other reasons, no one anointed IA to address the educational needs of the nation, and the Website has not been designed to meet specific

educational needs or even to focus only on books that are not commercially available. Section 110 of the Copyright Act, known as the TEACH Act, already grants teachers broad rights regarding the use of literary works, and book publishers, libraries and schools have on their own initiative sought to address educational needs directly during the health crisis while keeping in mind the needs of all stakeholders. In short, Defendant opportunistically seized upon the COVID-19 pandemic as an excuse to accomplish its long-desired goals while ignoring the law and harming the publishing ecosystem so critical to the world of books.

C. Defendant’s Infringement Causes Harm to Publishers

119. In all its different iterations (including under its “controlled digital lending” protocols), IA’s Website causes substantial harm to Plaintiffs, who produce and distribute books on behalf of themselves and their authors, to whom they pay royalties. Without the Publishers’ permission, IA and its Open Library business are using the unauthorized book scans to exploit existing markets (and potentially new related markets), causing authors and Plaintiffs substantial and irreparable harm. This includes, but is not limited to, the types of market harm outlined below.

120. First, a verbatim copy is a classic non-transformative use and a substitute for the original work. By operating the Website, IA competes directly with Publishers’ works in all formats (including, without limitation, print and digital) and market segments (including, without limitation, commercial, library, and school).

121. Second, by providing copies of digital books for free, IA devalues the book market. Consumers begin to view works as cheap and become increasingly unappreciative of what it takes to produce them and unwilling to pay fair value for them.

122. Relatedly, IA’s decision to provide free and full copies of Plaintiffs’ books, including the Works, to anybody who wants them interferes with Plaintiffs’ relationships with

customers and distributors, including libraries, who had already paid full value for them. The willingness of those distributors and customers to acquire digital formats in the future is diminished by the free distribution of Plaintiffs' books, including the Works, on the Website.

123. Third, public libraries are evolving to meet the changing needs of their patrons and trying hard to remain at the center of their communities. Defendant's Website undercuts public libraries by disintermediating them. Harm to and loss of community support for public libraries, in turn, hurts the Publishers, since these libraries pay for their books, which, in turn, hurts authors, who share in and depend upon compensation for their copyrights.

124. Fourth, the Website's .pdf and ePUB files are often of inferior quality. Authors expect their publishers to be guardians, ensuring the high-quality of their works as delivered to the marketplace. Websites like Open Library, thus, hurt Plaintiffs' relations with authors.

125. Likewise, inferior quality scans affect the Publishers' relationships with consumers. For example, in certain instances book scans from IA have surfaced on retailer websites such as Amazon as offerings by sponsored third-party vendors, leading to negative reviews on the product pages. When a consumer gives such a work a one-star review because of the scan quality, other consumers may only focus on the negative rating, even if it has nothing to do with the content of the work. More generally, consumers may be confused about whether the publisher has authorized the IA scan, leading to a negative impact on the publisher's goodwill.

126. Fifth, IA interferes with the author's and publisher's right to decide which works will be distributed in which format and at which time. For example, as noted above, some works or authors are ill-served by the conversion of print editions into digital works, either for commercial or artistic reasons. IA has appropriated to itself this right that belongs exclusively to the rightsholder.

127. Finally, Plaintiffs have legitimate fears regarding the security of their works both as stored by IA on its servers and subsequently publicly displayed or transmitted to users. For example, upon information and belief, IA has not developed and does not enforce sufficiently rigorous DRM protocols and related logistical systems that aggregators must employ to ensure that books appearing on their websites are not pirated or unlawfully infringed by users.

CLAIMS FOR RELIEF

FIRST CAUSE OF ACTION

Direct Copyright Infringement (17 U.S.C. § 101 *et seq.*)

128. Plaintiffs repeat and reallege each and every allegation above as if fully set forth herein.

129. The Works are original, creative, and copyrightable subject matter under the laws of the United States.

130. The copyrights in the Works are registered, and the Copyright Office has issued valid Certificates of Registration for the Works.

131. By its actions, alleged above, Defendant has infringed and will infringe the Publishers' copyrights in and to the Works by, *inter alia*, reproducing, distributing, publicly displaying, publicly performing, and making derivative works of the Works without any authorization or permission from Plaintiffs.

132. Each infringement of the rights of Plaintiffs in one of the Works constitutes a separate and distinct act of infringement.

133. Defendant's infringement of Plaintiffs' copyrights, including the Plaintiffs' Works, is willful.

134. Upon information and belief, as a direct and proximate result of its wrongful conduct, Defendant has and will obtain benefits, including, but not limited to, profits to which Defendant is not entitled.

135. As a direct and proximate result of Defendant's wrongful conduct, Plaintiffs have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Unless restrained by this Court, Defendant will cause further irreparable injury to Plaintiffs.

136. As a direct and proximate result of Defendant's infringement, Plaintiffs are entitled to recover statutory damages, pursuant to 17 U.S.C. § 504(c), with respect to each work infringed. Alternatively, at the election of Plaintiffs, pursuant to 17 U.S.C. § 504(b), Plaintiffs are further entitled to recover from Defendant the damages they have sustained and will sustain, as well as any gains, profits and advantages obtained by Defendant as a result of its acts of infringement alleged above. At present, the amount of such damages, gains, profits, and advantages cannot be fully ascertained by Plaintiffs.

137. Plaintiffs are entitled to recover their attorney's fees and costs.

SECOND CAUSE OF ACTION
Secondary Copyright Infringement (17 U.S.C. § 101 *et seq.*)

138. Plaintiffs repeat and reallege each and every allegation above as if fully set forth herein.

139. Although it is clear that IA is directly liable for copyright infringement, Plaintiffs also bring claims for secondary liability, in the alternative, to the extent IA attempts to evade responsibility for its own direct liability by blaming the conduct on others, such as users of the Website.

140. Defendant is secondarily liable under theories of contributory liability, inducement liability, and vicarious liability for the underlying reproduction, distribution, public display, and public performance of Plaintiffs' Works, as well as the making of infringing derivatives of Plaintiffs' Works.

141. Defendant is contributorily liable as it knows, or has reason to know, that Plaintiffs' Works are infringed each time Plaintiffs' Works are scanned, uploaded, downloaded, publicly displayed, or publicly performed in connection with the Website, and it has caused and/or materially contributed to those infringements.

142. Defendant is also secondarily liable under a theory of inducement. Defendant is responsible at every level for scanning copyrighted books, including the Works, uploading the scanned copies to its Website, and then distributing, displaying, and performing the works publicly. Defendant could take simple measures to prevent further infringement, such as by not scanning books under copyright, limiting the books on the Website to works in the public domain, or obtaining licenses to distribute ebooks. Instead, the infringement of copyrighted books, including Plaintiffs' Works, is a central focus of Defendant's business strategy. Defendant affirmatively encourages individuals to use the Website to infringe copyright, including by downloading digital copies of copyright protected books without a license, including the Works.

143. Further, Defendant is vicariously liable for copyright infringement because it has the right and ability to supervise the infringement of its users and possesses a financial interest in the infringement. As detailed above, Defendant has the right and ability to supervise the infringement of Plaintiffs' copyrighted books, including the Works. Defendant provides the sites and facilities on which the infringing activity occur. Defendant can stop allowing users of the

Website to download Plaintiffs' copyrighted books, including the Works. In brief, Defendant is in charge of the Website.

144. Defendant possesses an obvious and direct financial interest in the infringement. The uploading and downloading of increasing numbers of copyrighted works to the Open Library, including the Works, enhances the reputation of Defendant and its Website as a comprehensive source of free digital books. It draws more user registrations from people wishing to download from the Website and more donations of funds and books. The infringing files also increase the number of visitors to the Website, in turn increasing the numbers of people who view the link to purchase books at the Defendant's affiliate, Better World Books, which as it acquires more books, provides them to IA for scanning. Further, Defendant obtains a direct financial benefit by charging for the costs of scanning copyrighted books, which are then uploaded to the Website.

145. Each infringement of the rights of Plaintiffs in one of the Works constitutes a separate and distinct act of infringement.

146. Defendant's infringement of Plaintiffs' copyrights, including Plaintiffs' Works, is willful.

147. Upon information and belief, as a direct and proximate result of its wrongful conduct, Defendant has and will obtain benefits, including, but not limited to, profits to which Defendant is not entitled.

148. As a direct and proximate result of Defendant's wrongful conduct, Plaintiffs have been, and will continue to be, substantially and irreparably harmed in an amount not readily capable of determination. Unless restrained by this Court, Defendant will cause further irreparable injury to Plaintiffs.

149. As a direct and proximate result of Defendant's infringement, Plaintiffs are entitled to recover statutory damages, pursuant to 17 U.S.C. § 504(c), with respect to each work infringed. Alternatively, at the election of Plaintiffs, pursuant to 17 U.S.C. § 504(b), Plaintiffs are further entitled to recover from Defendant the damages they have sustained and will sustain, as well as any gains, profits, and advantages obtained by IA as a result of its acts of infringement alleged above. At present, the amount of such damages, gains, profits, and advantages cannot be fully ascertained by Plaintiffs.

150. Plaintiffs are entitled to recover their attorney's fees and costs.

PRAYER FOR RELIEF

WHEREFORE, Plaintiffs Hachette, HarperCollins, Penguin Random House, and Wiley, respectfully request judgment against Defendant Internet Archive as follows:

- A. Declaring that the practices of Internet Archive in connection with "Open Library" constitute willful copyright infringement;
- B. Issuing a preliminary and permanent injunction enjoining Internet Archive, and its agents, servants, employees, attorneys, successors and assigns, and all persons, firms and corporations acting in active concert or participation with it, from directly or indirectly reproducing, distributing, publicly displaying, creating derivative works, otherwise infringing, or causing, enabling, facilitating, encouraging, or inducing the reproduction, distribution, public display, creation of derivative works, or other infringement of, any of the respective copyrights owned or exclusively controlled, in whole or in part, by Plaintiffs, whether now in existence or hereinafter created, and ordering that all unlawful copies be destroyed;

- C. Entering judgment for Plaintiffs against Internet Archive for statutory damages in an amount based upon Internet Archive's willful acts of infringement of the Works, as alleged above, pursuant to the Copyright Act, 17 U.S.C. §§ 101, *et seq.*;
- D. Alternatively, ordering Internet Archive to render a full and complete accounting to Plaintiffs of Internet Archive's profits, gains, advantages, or the value of business opportunities received from the foregoing acts of infringement of the Works and entering judgment for Plaintiffs against Internet Archive for all damages suffered by Plaintiffs and for any profits or gain by Internet Archive attributable to the infringements alleged above of Plaintiffs' copyrights in amounts to be determined at trial;
- E. Awarding Plaintiffs the costs and disbursements of this action, including reasonable attorney's fees, pursuant to 17 U.S.C. § 505;
- F. Awarding Plaintiffs pre-judgment and post-judgment interest, to the fullest extent available, on the foregoing; and
- G. Granting such other further and different relief as this Court deems just and proper.

DEMAND FOR JURY TRIAL

Pursuant to Federal Rule of Civil Procedure 38, Plaintiffs demand a jury trial on all issues so triable in this action.

Dated: June 1, 2020
New York, New York

DAVIS WRIGHT TREMAINE LLP

/s/ Elizabeth A. McNamara

Elizabeth A. McNamara

Linda Steinman

John M. Browning
Meredith I. Santana
1251 Avenue of the Americas, 21st Floor
New York, NY 10020
Phone: (212) 489-8230
Email: lizmcnamara@dwt.com
lindasteinman@dwt.com
jackbrowning@dwt.com
meredithsantana@dwt.com

OPPENHEIM + ZEBRAK, LLP

Matthew J. Oppenheim
Scott A. Zebrak (*pro hac vice* motion forthcoming)
4530 Wisconsin Avenue, NW, 5th Floor
Washington, DC 20016
Phone: (202) 480-2999
Email: matt@oandzlaw.com
scott@oandzlaw.com

*Attorneys for Hachette Book Group, Inc.,
HarperCollins Publishers LLC, John Wiley & Sons,
Inc., and Penguin Random House LLC*