



21st Floor  
1251 Avenue of the Americas  
New York, NY 10020-1104

**Elizabeth A. McNamara**  
(212) 489-8230 tel  
(212) 489-8340 fax

[lizmcnamara@dwt.com](mailto:lizmcnamara@dwt.com)

August 12, 2021

**VIA ECF**

Hon Ona T. Wang  
United States District Court  
500 Pearl Street  
New York, NY 10007  
[Wang\\_NYSDChambers@nysd.uscourts.gov](mailto:Wang_NYSDChambers@nysd.uscourts.gov)

Re: *Hachette Book Group, Inc. et al. v. Internet Archive*, 1:20-CV-04160-JGK

Dear Magistrate Judge Wang:

We represent plaintiffs Hachette Book Group, Inc, HarperCollins Publishers LLC, John Wiley & Sons, Inc. and Penguin Random House LLC (“Plaintiffs”). Plaintiffs respectfully submit this response to the August 9, 2021 letter motion filed by Internet Archive (“IA”).

This lawsuit involves IA’s industrial-scale copying and online distribution of Plaintiffs’ and others’ in-copyright books. IA operates an illegal ebook distribution service that threatens to destroy the legal library ebook market and otherwise harm consumer book sales if left unchecked. IA accumulates millions of hard copies of copyrighted books as cheaply as possible, scans them, and distributes digital scans of the entire book without a license to anyone in the world who signs up to IA’s website. IA argues that this is legal fair use under a manufactured theory called “controlled digital lending.”

While Plaintiffs understand that IA has copied and distributed thousands of their books without authorization, this copyright infringement action identifies a representative subset of 127 books that are all currently licensed or sold as ebooks (the “Works in Suit”), ranging from classics by Zora Neale Hurston and J.D. Salinger to contemporary thrillers and romances.

Plaintiffs have produced a vast wealth of detailed sales and related financial data concerning the Works in Suit, totaling over 670,000 rows of data in Excel. Now, after the close of document discovery and on the eve of depositions, IA seeks to compel the production of “commercial performance data,” broken down by month, distribution channel, price and income, for *all* other books published by the Plaintiffs since 2011 – an undertaking that would involve a massive amount of data concerning more than 500,000 titles.

And IA makes this extraordinary demand in order to rifle through an enormous reservoir of highly proprietary data concerning books that are *not* the Works in Suit, all in an effort to somehow select “for each work in suit, one or more comparable books that were not available for digital lending” on IA’s system. Dkt. 47, 2. In other words, because the significant financial data already provided concerning the Works in Suit apparently does not support IA’s theory on

**DWT.COM**

Anchorage | Bellevue | Los Angeles | New York  
Portland | San Francisco | Seattle | Washington, D.C.

Hon. Ona T. Wang  
August 12, 2021  
Page 2

market harm, IA wants access to millions of data points concerning Plaintiffs' entire book catalogues. IA argues it is entitled to do this in order to see if any evidence might exist to support the inherently incredible theory that copying entire books and distributing them to any member of the public worldwide upon demand does not compete with Plaintiffs' sales of the same books. Even worse, IA's quest rests on the palpably false theory that it can quantify the harm caused by its infringement by comparing the sales of completely different books. Books are not interchangeable widgets and marketplace performance is driven by countless indeterminate and changing facts.

In short, Defendant's letter motion should be denied because the enormous and costly burden to Plaintiffs far outweighs the negligible value (if any) of the evidence sought, especially given the lack of legally relevant results that it will yield and the delays it will cause to the case.

### **THE DATA IA SEEKS IS NOT MATERIAL AND CREATES A MASSIVE BURDEN**

Plaintiffs' books are distributed to millions of readers around the world through many legal channels, including a thriving market for library ebook lending. Plaintiffs sell or license ebooks to library aggregators, who provide ebook platforms that millions of library patrons use to borrow ebooks on their devices for free, instantaneously and lawfully, from any location, subject to certain limitations.

The fourth factor of the fair use analysis examines the potential market harm from Defendant's action. IA argues that it needs sales data or comparables "to argue that the challenged library lending practice did not affect commercial performance" of the Works in Suit. This argument wrongfully cabins and misinterprets the fourth factor of the fair use analysis and evidences a false assumption about what drives book sales.

The fourth factor looks beyond lost sales and focuses instead on "the effect of the [infringing use] upon the *potential* market for or value of the copyrighted work." 17 U.S.C. §107(4). "Critically, it requires consideration of 'not only ... the market harm caused by the particular actions of the alleged infringer,' but also the market harm that would result from 'unrestricted and widespread conduct of the [same] sort.'" *Fox News Network v. TVEyes, Inc.*, 883 F.3d 169, 179-80 (2d Cir. 2018) (quoting *Campbell v. Acuff-Rose*, 510 U.S. 569, 590 (1994)).

Here, IA self-evidently harms Plaintiffs' existing markets by refusing to pay the customary fees that aggregators pay to distribute the same ebooks to library borrowers. The market harm compounds as others engage in the activity, destroying the vibrant market for licensed library ebook lending that currently exists. None of the requested data, or so-called "comparables," is necessary to or even related to this category of market harm.

Nor is the demanded sales data necessary to demonstrate potential market harm by showing that the "secondary use competes in the rightsholder's market as an effective substitute for the original." *Capital Records v. ReDigi Inc*, 910 F.3d 649, 662 (2d Cir. 2018). Over and over the Second Circuit has recognized that a non-transformative use of a work – as here – is likely to be "an effective substitute for the original." *Id.* at 662-63; *Authors Guild v. Google Inc.*,

Hon. Ona T. Wang  
August 12, 2021  
Page 3

804 F.3d 202, 223, 225-26 (2d Cir. 2015) (a book publisher would have a “strong” claim of market harm if their “claim were based on Google’s converting their books into a digitized form and making that digitized version accessible to the public”).

Even if sales data is the yardstick IA elects to focus on, to the extent that inquiry is relevant, Plaintiffs have already produced hugely detailed sales data relating to the Works in Suit for many years before and during IA’s infringement.<sup>1</sup> Seeking an additional “comparable” set of sales data for half of million other books is not only burdensome in the extreme, it is irrelevant.

*First*, IA’s argument that there is no burden in producing the data is preposterous. Contrary to IA’s suggestion that this information is available at the flip of a switch, the request would require Plaintiffs to pull masses of data from multiple databases, stitch it together and figure out a way to transfer enormous datasets in a way that they do not do in the ordinary course of business. An analyst for one Plaintiff estimated that it would take his entire team several weeks or months working full time to compile the requested information – which would amount to hundreds of millions, if not billions, of rows of data. The demand seeks access to data in a manner that lies far outside the bounds of how data is accessed, amassed or referenced by the Plaintiff publishers in the ordinary course of business and would involve incalculable employee time and significant resources. IA offers no precedent where a similar demand was allowed.<sup>2</sup>

*Second*, since books are not fungible widgets, the entire project rests on a false premise. There is no such thing as a “comparable book” – even if “comparable” is defined as some undefined period of sales data. Should *Catcher in the Rye* have similar sales to a best-selling cookbook, no one could plausibly contend the two works were “comparable.” Moreover, sales vary dramatically over different periods of time, depending on countless events (*e.g.*, an author dies; the book is made into a movie; it receives a great review or is selected by a state for its curriculum). No control group can begin to equate sales data.

In short, it is impossible to calculate the market harm of IA’s infringement based on the crude comparison IA proposes because there are innumerable reasons why one book sells more copies than another that have nothing to do with IA’s infringement. IA’s demand for this massive data rests on no coherent foundation. It should be denied.

Respectfully submitted,

/s/ Elizabeth A. McNamara

---

<sup>1</sup> IA demands monthly sales data for the Works in Suit. Plaintiffs have already produced monthly sales data from Hachette for 2020 and HarperCollins for a number of the works; the rest of the data reflects annual sales. To produce additional monthly sales is both burdensome and not necessary.

<sup>2</sup> IA cites the inapposite *Viacom Int’l Inc. v. YouTube Inc.*, 253 F.R.D. 256 (S.D.N.Y. 2008). Despite IA’s misleading parenthetical, that court’s order compelling YouTube, the *defendant*, to produce a minimal quantity of data concerning certain non-infringing videos to support *plaintiffs’* infringement is irrelevant to IA’s demand for billions of rows of data for every work published by Plaintiffs in the last decade.

Hon. Ona T. Wang  
August 12, 2021  
Page 4

cc: Counsel of Record (via ECF)