



Joseph C. Gratz
415-376-6407 (direct)
415-362-6666 (main)
jgratz@durietangri.com

August 9, 2021

VIA ECF

Hon. John G. Koeltl
U.S. District Judge
Daniel Patrick Moynihan
United States Courthouse
Courtroom 14A
500 Pearl St.
New York, NY 10007-1312

Re: *Hachette Book Group, Inc. et al. v. Internet Archive*
Case No. 1:20-CV-04160-JGK

Your Honor:

Pursuant to Local Civil Rule 37.2, Defendant Internet Archive respectfully requests a pre-motion discovery conference regarding a motion to compel the production of information regarding the commercial performance of books published by Plaintiffs.

In the above-captioned lawsuit, Plaintiffs contend that the Internet Archive infringed Plaintiffs' copyrights by the non-profit digital lending of library books. The Internet Archive maintains that the challenged lending constitutes fair use under 17 U.S.C. § 107.

In considering fair use, one factor courts consider is "the effect of the use upon the potential market for or value of the copyrighted work." Plaintiffs claim that the Internet Archive's digital library lending has a negative effect on the market for or value of the works. The Internet Archive disagrees, and wishes to bring forward evidence showing that lending had little or no effect on the commercial performance of the books being lent, compared to books that were not lent. "Since fair use is an affirmative defense, its proponent would have difficulty carrying the burden of demonstrating fair use without favorable evidence about relevant markets." *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 590 (1994). The Internet Archive submits this letter-motion in order to gather that evidence.

Specifically, in order to show that lending had little or no effect on commercial performance, the Internet Archive wishes to compare the commercial performance of books that were available for digital lending with books that were not available for digital lending. The Internet Archive has requested data

Hon. John G. Koeltl
August 9, 2021
Page 2

about the commercial performance¹ of all of Plaintiffs' books, broken down by month, since 2011. Plaintiffs have provided some commercial performance data, but (1) Plaintiffs have refused to provide data about books other than the works-in-suit, and (2) the data they have provided is not broken down by month. For reasons explained below, Plaintiffs can and should produce this data, and the Internet Archive asks that this Court compel them to do so.

A. Commercial performance data about books other than the works-in-suit is necessary to understand whether lending had any effect on commercial performance.

In order to argue that the challenged library lending practice did not affect commercial performance, one needs commercial performance data not only for the books that were lent out, but also of other books that were not loaned. Without that data, the Internet Archive has nothing to compare. *See, e.g., Viacom Int'l Inc. v. Youtube Inc.*, 253 F.R.D. 256, 261 (S.D.N.Y. 2008) (granting motion to compel production of data about all videos not accused of infringement because such data was necessary "to compare the attractiveness of allegedly infringing videos with that of non-infringing videos.").

Plaintiffs object that data about other books would be irrelevant. They have argued that, because there are too many other factors that could affect their commercial performance, the data won't show whether the Internet Archive's digital library lending affected commercial performance. We have no doubt that Plaintiffs will press that line of argument in cross-examination of the Internet Archive's witnesses, but such estimates are a necessary part of litigation about alleged copyright infringement. *See, e.g., Davis v. The Gap, Inc.*, 246 F.3d 152, 166–67 (2d Cir. 2001) ("Many of the accepted methods of calculating copyright damages require the court to make uncertain estimates in the realm contrary to fact."). Plaintiffs' line of argument provides no justification for withholding the data necessary to perform the comparison in the first place.

Plaintiffs also object that producing data about *all* other books would be unduly burdensome, because there are only 127 works-in-suit. The Internet Archive agrees that the comparison does not require each and every book; instead, it requires, for each work-in-suit, one or more comparable books that were not available for digital lending at the same time as that work-in-suit. One way to identify potentially comparable books would be to identify books whose commercial performance was very similar before either book was made available for digital lending. But Plaintiffs, who are in possession of the data one would need to do that analysis, have declined to identify books they regard as comparable—because, as discussed above, they take the position that no book is comparable to any other book. Given this refusal, Plaintiffs must produce data about all books, so that the Internet Archive can identify books it

¹The types of commercial performance data the Internet Archive seeks are: the number of physical copies embodying works by distribution channel, the prices for those physical copies, the number of ebooks embodying works by distribution channel, the number of ebook loans divided by distribution channel, prices for any transaction related to ebooks, income from sales of physical books by distribution channel, and income from ebook transactions by distribution channel. RFPs 20-22, 60-67.

Hon. John G. Koeltl
August 9, 2021
Page 3

regards as comparable, and the parties can then debate, on a level playing field, whether such books are or are not comparable.

Nor is there any particular burden in retrieving the information requested. This is commercial data stored in databases, indexed by book. Plaintiffs were able to provide data about the works-in-suit by accessing such databases; this motion simply seeks the result of querying the same systems for a larger set of books.

B. Plaintiffs must produce *monthly* commercial performance data, not just annual data.

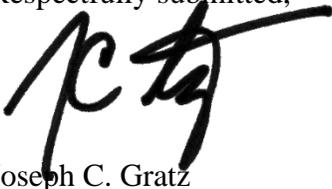
Hachette, Penguin Random House, HarperCollins, and Wiley have produced annual commercial performance data for the works-in-suit. They do not contend that monthly data does not exist; they contend only that producing monthly data would be more difficult, so they have not produced it. Indeed, Hachette produced monthly data for 2020, but not for the other requested years. And, of course, as discussed above, Plaintiffs have not produced any data for books that are not works-in-suit.

Monthly data is necessary because each book began to be available for digital lending on a particular known date, so it makes sense to compare the commercial performance *before* that particular date with the commercial performance *after* that date. Purely annual data does not provide enough detail, particularly because sales of a particular book change so drastically within a year. (*The New York Times* publishes its bestseller lists weekly, after all, for this reason.)

Monthly data is also necessary because one issue in the case is whether commercial performance was affected differently during the National Emergency Library. This was the approximately three-month period, during the first wave of shelter-in-place orders that shuttered schools and libraries, when the Internet Archive made books available for lending without imposing the strict one-to-one “paper-copies-owned-to-digital-copies-loaned” ratio it imposes during normal operations. Without monthly data, there is no way to tell whether that short-lived emergency practice affected commercial performance.

For the reasons explained above, and pursuant to Local Rule 37.2 and Your Honor’s Individual Practices Paragraphs 1F and 2B, Defendant respectfully requests a pre-motion conference with the Court to discuss: Plaintiffs’ production of monthly commercial performance data since 2011 for all of their books in print during that period.

Respectfully submitted,



Joseph C. Gratz

JCG:co