



Scott A. Zebrak
4530 Wisconsin Avenue NW, 5th Floor
Washington, DC 20016
202.450.3758 | scott@oandzlaw.com

November 19, 2021

Via ECF

Hon. John G. Koeltl
United States District Judge
United States Courthouse
500 Pearl Street
New York, NY 10007

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

Dear Judge Koeltl:

We represent plaintiffs Hachette Book Group, Inc, HarperCollins Publishers LLC, John Wiley & Sons, Inc., and Penguin Random House LLC (Plaintiffs”) in this case, in which fact discovery closes on December 17, 2021. We respectfully submit this letter to request a Local Rule 37.2 Conference regarding a motion to compel Defendant Internet Archive (“IA”) to produce documents. Plaintiffs have patiently afforded IA opportunity after opportunity to resolve this discovery dispute. But IA is stonewalling, hoping to run out the clock on discovery.

Background

Plaintiffs are among the leading book publishers in this country. Plaintiffs commenced this copyright infringement action against IA on June 1, 2020.

IA is a mass infringer. IA engages in industrial scale scanning of copyrighted print books; uploads the resulting eBooks to its Internet-connected servers; and distributes these eBooks in their entirety from its website for periods of up to fourteen days—all under a contrived theory called “Controlled Digital Lending.” IA’s unlawful digitization and online distribution of millions of copyrighted books occurs without any license from, or payment to, the rightsholders.

Discovery has proceeded slowly, in large part because IA delayed a substantial portion of its document production until toward the end of the extended deadline for document discovery. Plaintiffs promptly reviewed that eleventh-hour production and then proceeded into depositions. It has become apparent that IA did not produce several categories of key documents. Because of all this, the remaining depositions of IA, including among its most key witnesses, are now not scheduled until December 3, 8, 9 and 17.

The Documents to be Produced

Plaintiffs met and conferred with IA multiple times, including in a lengthy call on November 9, 2021, as well as via numerous emails before and afterwards, but to no avail. IA has not disputed the relevance of the documents. Instead, IA is simply stonewalling, prejudicing Plaintiffs in depositions and case development. For the categories of documents below, Plaintiffs have waited several weeks or more for IA to respond and followed up several times.

Category 1: Plaintiffs informed IA that its production does not appear to include technical specifications, policies, procedures, or source code for IA designating items (books) into its digital “lending library” (a/k/a “Books to Borrow” or “inlibrary” collection). Plaintiffs asked IA to indicate if its document production includes these materials and to provide some direction on what to look for and where, or, alternatively, to obtain and produce such documents. IA has not provided a substantive response. The documents at issue are responsive to multiple requests, including RFP Nos. 10, 11, 16, 17, 18, 28 and 80.

Category 2: Plaintiffs informed IA that its production does not appear to include policies or procedures for how IA conducts overlap analyses between IA’s digital holdings and a library’s physical holdings and then implements the results, including to adjust lending counts as to the purported overlap. For instance, among many other open issues, IA’s comparison produces both “direct matches” and “similar matches.” IA must produce its documents explaining what IA considers is related enough to be a similar match, and how IA treats lending counts when its data analysis (not review of actual books) finds “similar matches.” The documents at issue are responsive to multiple requests, including RFP Nos. 7, 14, 17, 18, 77, 80 and 81.

Category 3: IA downloaded copyrighted books, including at least one work in suit, from a patently unlawful offshore pirate website called Library Genesis, or “LibGen,” which has been the subject of injunctive relief orders from multiple courts around the world. IA then uploaded the digital files to its website, placing them within something it called the “Internet Books” (or librarygenesis) collection. Plaintiffs have requested IA produce certain technical files for the works in suit that IA obtained from LibGen and then put onto its website in the Internet Books collection. The requested materials would enable Plaintiffs to determine the extent to which IA then distributed further copies to users of its website, including in what context(s) and under what restrictions, if any. The documents at issue are responsive to multiple requests, including RFP Nos. 3, 4, 5, 6, 7 and 9.

Category 4: Since as far back as September 29, Plaintiffs repeatedly asked IA to direct Plaintiffs to documents, including any supporting studies, analyses or data, in IA’s production that support its contention that the digitized copies of copyrighted books that IA provides to users of its website are typically only borrowed for a short period of time. The documents are responsive to multiple requests, including RFP Nos. 5, 17, 26, 63, 64 and 65. In response, IA directed Plaintiffs in two directions. First, IA referred Plaintiffs to certain data sets for the works in suit. But the data sets that IA produced, in the form of “CSV” files, are not sufficient to calculate the length of specific loans—at least not where IA has multiple concurrent users of copies of a single digitized print book. As to other data sets that IA produced, in the form of “JSON” files, certain data in the files has been cryptographically scrambled. Accordingly,

Plaintiffs asked IA whether, in these JSON files or elsewhere, IA has data it can produce to show when each loan was made and when each loan was ended. IA has not responded. Second, IA referred Plaintiffs to a handwritten, undated document (bates number INTARC00151679) that analyzes a single day of activity. Despite multiple requests from Plaintiffs, however, IA has not indicated if it will produce the underlying and related data referenced, relied upon, and/or analyzed in INTARC00151679. Finally, nor has IA confirmed if there are any other documents that supposedly support its contention as to the length of borrows.

Category 5: Plaintiffs requested that IA produce data that reflects how users of IA’s website engage with the digitized copies of copyrighted books that IA provides. IA admitted that it copied, uploaded, and distributed each work-in-suit. However, IA contends that it can only provide the overall number of loans per title, and that server log information as to how frequently the user reads the book in the Internet Archive book reader interface, or downloads a PDF or ePub copy for offline reading, is not reasonably accessible to IA without disproportionate burden and cost. At this time, Plaintiffs do not ask the Court to compel IA to produce that data. However, there appears to be a different category of data that IA can and should produce, which concerns use of a “listen” feature whereby users of IA’s website can have the digitized copy of the print book read to them, akin to an audio book. The documents at issue are responsive to multiple requests, including RFP Nos. 5, 6 and 17. In the parties’ conferral, IA did not assert that the data concerning audio streams is inaccessible due to unreasonable burden or cost.

Category 6: Plaintiffs requested that IA produce emails or other communications from users who complained when IA switched from standard 14-day loans to initial 1-hour loans for eBooks where IA has only one physical copy of the book. The documents at issue are responsive to multiple requests, including RFP Nos. 11, 17, 33, 63, 64 and 65. Indeed, the requested documents will show that users were upset when their reading of a book or series was interfered with by the new policy, evidencing that the books are indeed read in full by users. IA has not confirmed whether it will produce the additional documents or when.

Category 7: During depositions of two IA witnesses, Lila Bailey and Jacques Cressaty, Plaintiffs learned that IA had not produced a variety of relevant financial documents and policy documents regarding “Controlled Digital Lending.” They are responsive to multiple requests, including RFP Nos. 22, 23, 24, 25, 54, 55 and RFP Nos. 17, 32, 38, respectively. IA has not confirmed when, or if, it will produce the documents. Similarly, an IA witness, Andrea Mills, testified that certain documents addressing IA’s current digitization process may not be within IA’s existing production. Plaintiffs asked for Internet Archive to produce them, but IA has not confirmed if it will supplement its production.

Plaintiffs thank the Court for its time and consideration on this matter.

Respectfully submitted,

/s/ Scott A. Zebrak

Scott A. Zebrak

cc: Counsel of Record (via ECF)