**Submission of MPA, N/MA, and RIAA**
**Class 3(a) & (b): Motion Pictures and Literary Works – Text and Data Mining**

UNITED STATES COPYRIGHT OFFICE

# Long Comment Regarding a Proposed Exemption Under 17 U.S.C. § 1201

**[ ] Check here if multimedia evidence is being provided in connection with this comment.**

## ITEM A. COMMENTER INFORMATION

**The Motion Picture Association, Inc.** ("MPA") is a trade association representing some of the world's largest producers and distributors of motion pictures and other audiovisual entertainment for viewing in theaters, on prerecorded media, over broadcast TV, cable and satellite services, and on the internet. The MPA's members are Netflix Studios, LLC, Paramount Pictures Corporation, Sony Pictures Entertainment Inc., Universal City Studios LLC, Walt Disney Studios Motion Pictures, and Warner Bros. Entertainment Inc.

**The News/Media Alliance** ("N/MA") represents over 2,200 publishers in the U.S. and internationally, ranging from the largest news and magazine publishers to hyperlocal newspapers, and from digital-only outlets to papers who have printed news since before the Constitutional Convention. Its members produce quality journalistic and creative content that accounts for nearly 90 percent of daily newspaper circulation in the U.S., over 500 individual magazine brands, and dozens of digital-only properties.

**The Recording Industry Association of America, Inc.** ("RIAA") is a nonprofit trade organization that supports and promotes the creative and financial vitality of recorded music and the people and companies that create it in the United States. RIAA's several hundred members—ranging from major American music companies with global reach to artist-owned labels and small businesses—make up the world's most vibrant and innovative music community. RIAA's members create, manufacture, and/or distribute the majority of all legitimate recorded music produced and sold in the United States. In supporting its members, RIAA works to protect the intellectual property and First Amendment rights of artists and music labels.

Represented By:
Robert H. Rotstein (rhr@msk.com)
James Berkley (jdb@msk.com)
Stacey Chuvaieva (stc@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
2049 Century Park East, 18th Floor
Los Angeles, CA 90067
301-312-2000

J. Matthew Williams (mxw@msk.com)
Lucy Holmes Plovnick (lhp@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
1818 N Street, NW, 7th Floor
Washington, D.C. 20036
202-355-7904

## ITEM B. PROPOSED CLASS ADDRESSED

Proposed Class 3(a) & (b): Motion Pictures and Literary Works – Text and Data Mining

## ITEM C. OVERVIEW

In the 2021 proceeding, current Class 3 proponents Authors Alliance, the American Association of University Professors and the Library Copyright Alliance sought "text and data mining" ("TDM") exemptions that are now codified in 37 C.F.R. § 201.40(b)(4) and (5). We did not oppose renewal of those exemptions, and the Office has already said it intends to recommend renewing them.[1] The proponents now argue that they gave up too much as the 2021 proceeding progressed, and that instead of merely retaining a limitation *they* proposed in 2021, the Register should now recommend much broader exemptions that would turn every researcher with an interest in "text and data mining techniques" into a distributor of databases of unprotected motion pictures and literary works without effective security measures. The proponents fail to provide any meaningful justification for such an expansion. To the contrary, the limitation they proposed in 2021 is appropriate and should be retained.

In the last proceeding, the TDM proponents initially proposed boundless exemptions that seemingly would have permitted circumvention of the technological measures used to protect valuable motion pictures and literary works by anyone for any purpose, so long as it was done "to deploy text and data mining techniques."[2] Such exemptions would have covered commercial and other infringing uses of the works involved and would have posed grave security risks related to the dissemination of such works.

Responding to criticisms as the 2021 proceeding continued, the proponents did not even attempt to justify the full scope of the exemptions they originally proposed, but instead agreed "to a narrower exemption provided that the scope of the exemption remain[ed] broad enough to permit their teaching and research to go forward."[3] The proponents proffered such a narrower exemption, which they stated "addresses concerns raised by opponents while allowing circumvention to enable the research activities described in the petition."[4] Specifically, their revised exemptions provided that researchers relying on the exemptions would "limit access to the corpus of circumvented works only to other researchers affiliated with qualifying institutions

---

[1] Exemptions To Permit Circumvention of Access Controls on Copyrighted Works: Notice of Proposed Rulemaking, 88 Fed. Reg. 72,013, 72,018-19 (Oct. 19, 2023).

[2] Authors Alliance *et al*., Class 7(a) & 7(b) Initial Comment at 4 (Dec. 14, 2020), https://www.copyright.gov/1201/2021/comments/Class%2007a%20and%2007b_InitialComments_Authors%20Alliance,%20American%20Association%20of%20University%20Professors,%20and%20Library%20Copyright%20Alliance.pdf ("AA 2021 Initial Comment").

[3] Authors Alliance *et al*., Class 7(a) & 7(b) Reply Comment at 5 (Mar. 9, 2021), https://www.copyright.gov/1201/2021/comments/reply/Class%207(a)%20and%207(b)_Reply_Authors%20Alliance,%20American%20Association%20of%20University%20Professors,%20Library%20Copyright%20Alliance.pdf ("AA 2021 Reply"); *see also id.* at 29 ("Petitioners do not object to accommodating legitimate concerns to the extent that the core purpose is preserved.").

[4] *Id*. at 6.

*for purposes of collaboration or the replication and verification of research findings*."[5] Notably, the proponents acknowledged that this "more narrowly tailored exemption" would "still enable[] the exemption's core purpose."[6]

Ultimately, the Register recommended exemptions that, as relevant here, hew closely to the language that the proponents maintained would preserve the "core purpose" of their initial proposals, with language limiting access to "researchers affiliated with other institutions of higher education solely *for purposes of collaboration or replication of the research*."[7]

Curiously, the proponents' primary argument is now that a word *they* actually proposed in 2021—"collaboration"—is so ambiguous that it "prevents researchers and teachers from effectively using the current exemption."[8] But, that wasn't what they said in 2021 when they wholeheartedly embraced the term as a way to enable the research activities described in their petition. For example, they wrote:

> Petitioners do not object to narrowing their proposal so long as the core purposes of advancing research and teaching are preserved. For example, petitioners would not object to excluding computer programs and limiting eligible works to those lawfully obtained by the researcher or their affiliated institution; limiting beneficiaries of the exemption to researchers affiliated with nonprofit libraries, archives, museums, or institutions of higher education; limiting the purpose of the exemption to scholarly research and teaching; *limiting access only to other researchers for project collaboration and verification of research results*; and requiring the use of reasonable security measures.[9]
>
> . . .
>
> Researchers at qualifying institutions are well-situated to appropriately manage and secure their research corpora in a way

[5] *Id*. (emphasis added).

[6] *Id*.

[7] SECTION 1201 RULEMAKING: EIGHTH TRIENNIAL PROCEEDING TO DETERMINE EXEMPTIONS TO THE PROHIBITION ON CIRCUMVENTION, RECOMMENDATION OF THE REGISTER OF COPYRIGHTS, 121-24 (Oct. 2021), https://cdn.loc.gov/copyright/1201/2021/2021_Section_1201_Registers_Recommendation.pdf ("2021 Rec."); *accord* 37 C.F.R. § 201.40(b)(4)(i)(D), (5)(i)(D).

[8] Authors Alliance *et al.*, Class 3 Long Comment at 8-10 (Dec. 21, 2023), https://www.copyright.gov/1201/2024/comments/Class%203(a)%20and%203(b)%20-%20Initial%20Comments%20-%20Authors%20Alliance,%20Library%20Copyright%20Alliance,%20and%20Am.%20Association%20of%20Uni.%20Professors.pdf ("AA 2023 Comment").

[9] *See* AA 2021 Reply at 3 (emphasis added).

> that limits access to *only collaborators and peer reviewers seeking to validate research findings*.[10]
>
> . . .
>
> Petitioners are willing to expressly limit access to research corpora to other researchers *for the limited purposes of collaboration on specific research projects and verification of research results*.[11]
>
> . . .
>
> [P]roponents need access to the entirety of the works they place in their corpora but *do not seek to provide access to others but for the purposes of collaboration or the replication and verification of research findings*. . . . And, again, the exemption does not contemplate providing access to the corpora to anyone but other researchers, and then *only for purposes of collaboration or the replication and verification of research findings*.[12]
>
> . . .
>
> To be clear, the exemption does not contemplate redistribution of the research corpus and other researchers' access to it is limited to *purposes of collaboration or the replication and verification of research findings*.[13]

These limitations were important to the case the proponents made for their proposed exemptions. For example, they stressed how UC Berkeley and other "key research hubs" have data management professionals to "secur[e] sensitive types of data, creat[e] access controls for researchers and collaborators, and disseminat[e] research results responsibly."[14] They also promised that "institutions have data storage options that can restrict data access to specific accounts of approved researchers affiliated with the institutions and apply granular controls to data depending on the level of sensitivity."[15]

The proponents do not identify any legal claims that have been made against researchers, or any other developments in the last three years, that might cause researchers to be "stymied" by "uncertainty" concerning the commonplace concept of research collaboration as embodied in the

---

[10] *Id*. at 13-14.

[11] *Id.* at 14 (emphasis added).

[12] *Id.* at 21 (emphasis added).

[13] *Id*. at 22 (emphasis added).

[14] *Id*. at 12-13.

[15] *Id*. at 13.

language proposed by the proponents in the last proceeding.[16] It appears that the real issue here is that the proponents wish to slowly chip away at copyright protection by enabling more "sharing" without any change in circumstances.[17] Thus, without any attempt to clear up the purported ambiguity, they seek to expand the current exemptions by tacking on to the current access limitation the following additional italicized permission:

> The institution uses effective security measures to prevent further dissemination or downloading of motion pictures in the corpus, and to limit access to only the persons identified in paragraph [(b)(4)(i)(A) or (b)(5)(i)(A)] of this section or to researchers affiliated with other institutions of higher education solely for purposes of collaboration or replication of the research*; or for the purposes of conducting independent text and data mining research and teaching, where those researchers are in compliance with this exemption.*[18]

What is ambiguous is the scope of this proposed expansion, not the term "collaboration"—a term that the proponents themselves used and relied on in assuring that the exemptions they proposed in 2021 would be narrowly applied.

Their proposed new language would dramatically enlarge the scope of the exemptions adopted in 2021—which the proponents proposed in their 2021 reply comments—although it isn't clear how far they would like it to extend or how it would possibly work in practice. For example, the reference to recipient researchers "in compliance with this exemption" apparently is intended to mean that the recipients would need to be "affiliated with a nonprofit institution of higher education" that has "lawfully acquired" or "licensed . . . without a time limitation on access" all of the works involved.[19] It also appears that this language would permit recipient researchers to redistribute corpora they had received from others "in compliance with this exemption," setting up a system where databases of unprotected motion pictures and literary works would be freely circulating among university researchers. However, while talking at length about how convenient it would be for researchers to be able to access vast databases of copyrighted works that have been stripped of their technological protection measures, the proponents' comments have nothing at all to say about how those requirements could ever be enforced. This proposal would compromise the security of large databases of valuable copyrighted works and permit a wide range of potentially infringing uses.

The Register should not give in to the proponents' strategy of seeking to diminish copyright protections bit by bit in each triennial cycle. The proponents fail to establish that all of the conduct covered by their proposed exemption is likely noninfringing or that the Section

---

[16] *See, e.g.*, AA 2023 Comment at 8.

[17] *See, e.g., id.* at 5, 8-12, 15, 17-18, 22, 28, 30, 31-32.

[18] *Id*. at 5-6.

[19] 37 C.F.R. § 201.40(b)(4)(i)(A), (B), (5)(i)(A), (B); *see* AA 2023 Comment at 5 ("that institution must itself own lawfully acquired copies of the underlying works"), 22 ("the beneficiary intuitions [sic] would still be required to obtain lawful copies of the underlying works or licenses without a time limitation on access").

1201(a)(1)(C) factors, as properly construed, support granting the proposed exemption. We thus oppose the requested modification to the current TDM exemptions.[20]

## ITEM D. TECHNOLOGICAL PROTECTION MEASURE(S) AND METHOD(S) OF CIRCUMVENTION

The access controls at issue include every access control applied to ebooks and a wide range of access controls used to protect motion pictures. Specifically, the proposed expanded exemption would extend to not only the Content Scramble System used on DVDs and the Advanced Access Content System use on Blu-ray discs, but also various technological protection measures used on digital download services, and (potentially) cable and satellite set-top boxes and video game consoles that receive motion picture downloads.[21] Many of these access controls enforce terms of use that allow for lower cost, temporary access and do not allow for the retention of permanent reproductions. These are precisely the kinds of access controls Congress intended to incentivize and protect when enacting Section 1201, because they increase the availability of motion pictures and literary works at affordable prices through access-based business models.[22]

The methods of circumvention at issue are any and all methods available. The proponents have identified no limitations.

## ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGING USES

As we acknowledged in the last proceeding, TDM "has the potential to benefit copyright owners, researchers, teachers, technology companies, and other users of copyrighted works."[23] For that reason, we did not oppose renewal of the existing TDM exemptions, and the Office has already said it intends to recommend renewing them.[24] However, our non-opposition to renewal of the current exemptions should not be taken as an indication that they are entirely satisfied with them. For example, as described below, the security provisions in 37 C.F.R. § 201.40(b)(4)(ii)(B) are not working effectively to give copyright owners confidence that their valuable copyrighted

---

[20] Our comments focus on motion pictures (Class 3(a)), because we understand that the Association of American Publishers is filing separate comments to address issues concerning literary works (Class 3(b). However, the issues are similar, and we do not support either proposed exemption.

[21] *See* 37 C.F.R. § 201.40(b)(4)(i) (beginning the definition of the works involved with reference to "[m]otion pictures . . . on a DVD protected by the Content Scramble System, on a Blu-ray disc protected by the Advanced Access Content System, or made available for digital download"). Given that the Register and Librarian have previously denied proposed exemptions for circumventing access controls on video game consoles, the proposals are best interpreted to exclude these devices from their scope even though the devices are used to receive downloads of motion pictures. If any expansion is granted, it should make these implied limitations express.

[22] *See, e.g.*, U.S. COPYRIGHT OFFICE, SECTION 1201 OF TITLE 17 at 9-10 (2017) https://www.copyright.gov/policy/1201/section-1201-full-report.pdf ("Section 1201 Report") ("By providing legal protection for access controls, Congress hoped to encourage copyright owners to make their works available to consumers through flexible and cost-effective online platforms" including ones that "allow access during a limited time period").

[23] Joint Creators and Copyright Owners, Class 7(a) & (b) Long Comment at 2 (Feb. 9, 2021), https://www.copyright.gov/1201/2021/comments/opposition/Class_7a%20and%207b_Opp'n_Joint%20Creators%20and%20Copyright%20Owners.pdf.

[24] 88 Fed. Reg. at 72,018-19.

works are being secured appropriately and consistent with the measures used to protect academic institutions' "own highly confidential information."[25]

The fact remains that TDM continues to raise complex concerns, in ways that are increasingly relevant and accentuated by the current proposal to broaden the exemption. The Register should reject the proposed changes constituting Class 3 because (1) enabling every university researcher with an interest in TDM to distribute vast databases of copyrighted works that have been stripped of their technological protection measures presents grave security and infringement risks; (2) the proponents have not established that the additional activity that would be enabled by the broadened exemption would be a fair use; and (3) this is neither the place nor the time to break new legal ground in deciding that question.

### 1. The proposal presents grave security and infringement risks.

The proponents' proposed new language would dramatically enlarge the scope of the exemption adopted in 2021 by permitting distribution and apparently redistribution of corpora containing unlimited numbers of copyrighted works across any number of institutions for purposes entirely independent of the ones for which they were created. Analysis of the proposal must focus on that activity rather than the propriety of TDM research in general.[26] We have significant concerns that the broad scope of the proposed exemption includes and will enable infringement of valuable copyrighted works. A non-exhaustive list of problems raised by the proposed exemption includes the ones identified in the paragraphs below.

**Market Substitution.** As noted above, in the 2021 proceeding, the proponents stated that "*the exemption does not contemplate redistribution of the research corpus* and other researchers' *access to it* is limited to purposes of collaboration or the replication and verification of research findings."[27] The proponents' proposed expansive new language, as explained in the proponents' comments, would seem to supplant the limitations of the existing exemptions—which were intended to allow only *access* to corpora, and only in limited circumstances—into a broad distribution entitlement.[28] Doing so would for the first time not only allow researchers that do not have a close working relationship with the creators of a research database of unsecured works to gain access to a copy of that database, but also for the first time apparently allow those falling within the exemption to obtain their own copy and make and redistribute further copies of that database, not only for purposes of research, but also for purposes of *teaching*.

It is not clear from the proposed regulatory language whether the teaching referenced is supposed to be teaching about TDM or teaching in general. If the latter, use of motion pictures and literary

[25] 37 C.F.R. § 201.40(b)(4)(ii)(B).

[26] *See* NPRM, 88 Fed. Reg. at 72,026 ("In cases where a class proposes to expand an existing exemption, participants should focus their comments on the legal and evidentiary bases for modifying the exemption, rather than the underlying exemption.").

[27] *See* AA 2021 Reply at 22 (emphasis added).

[28] *See, e.g.*, AA 2023 Comment at 5 ("allow sharing with researchers affiliated with different nonprofit institutions"), 17 ("requests from other researchers for corpora"; "[e]nabling broader sharing"), 18 ("permitting researchers to share corpora with researchers at other institutions").

works for any teaching purposes (including showing an entire motion picture in a class on cinema) clearly includes uses that do not constitute fair use.[29] Either way, the effect, and perhaps even the goal, seems to be to put massive databases of unsecured motion pictures and literary works into relatively free circulation among educational institutions. This proposal would open up access to a much larger number of people, increasing the risk that decrypted motion pictures and literary works would be used for expressive or otherwise substitutive purposes. Such a use requires a justification that is absent here.[30]

**Ownership**. The proponents state that under the proposed new expansion, independent researchers may acquire databases compiled by other institutions only if their own institution owns lawfully acquired underlying copies of the copyrighted works in those databases.[31] However, the proponents do not give any hint of how creators of corpora (much less users several steps down the distribution chain) could possibly verify that information, especially where a corpus contains hundreds of thousands of unprotected motion pictures or literary works.[32] It seems impossible. And there is no guarantee that a downstream database recipient's institution would ever provide ownership information to copyright owners. In contrast, institutions that create their own databases necessarily do so by decrypting and reproducing copies they possess—and more likely own and have acquired lawfully. Thus, there is a heightened and significant risk that downstream recipients of a corpus would receive and access unprotected copyrighted works that their institution does *not* own, and that copyright owners would be unable to detect and investigate noncompliance. Even if the ownership requirement were workable (and it is not), such a limitation would fail to ensure that the expanded uses are fair. In other situations where ownership verification has been relied on to justify "space shifting," courts and the Librarian have rejected fair use arguments.[33]

---

[29] Although Section 110(1) permits certain public performances of complete motion pictures and literary works *"in a classroom or similar place devoted to instruction"* (emphasis added), without obtaining licenses, it does not allow those performances to be generated from unauthorized copies. Section 110(1) also does not permit reproductions of copyrighted works.

[30] *See Andy Warhol Found. for the Visual Arts, Inc. v. Goldsmith*, 598 U.S. 508, 532 (2023) ("An independent justification . . . is particularly relevant to assessing fair use . . . where wide dissemination of a secondary work would otherwise run the risk of substitution for the original or licensed derivatives of it.").

[31] AA 2023 Comment at 5, 22.

[32] In the prior rulemaking proceedings, the proponents emphasized that they were agreeable to limiting the scope of the exemption such that each institution must create its own database of motion pictures or literary works, specific to each research project, using copies that institution itself lawfully acquired. *See, e.g.,* AA 2021 Reply at 6, 22. They ultimately disclaimed a desire for institutions to share databases or to enable linking of multiple databases. *See id.*at 5-6, 22-23, 29. The proposed new language conflicts with that prior position.

[33] *E.g., Hachette Book Grp., Inc. v. Internet Archive*, 664 F. Supp. 3d 370, 380 (S.D.N.Y. 2023) ("Other cases consistently have held that the first fair use factor weighs against infringers who do nothing more than 'change[ ] the format' of a pre-existing work.…"); *Capitol Recs., LLC v. ReDigi Inc*., 910 F.3d 649, 663-64 (2d Cir. 2018); *UMG Recordings, Inc. v. MP3.Com, Inc*., 92 F. Supp. 2d 349, 350 (S.D.N.Y. 2000) ("Thus, although defendant seeks to portray its service as the 'functional equivalent' of storing its subscribers' CDs, in actuality defendant is re-playing for the subscribers converted versions of the recordings it copied, without authorization, from plaintiffs' copyrighted CDs."). Indeed, Courts and the Copyright Office have repeatedly concluded that space shifting and format shifting are not fair uses. *See, e.g., Disney Enters., Inc. v. VidAngel, Inc.,* 869 F.3d 848, 862 (9th Cir. 2017) ("The reported decisions unanimously reject the view that space-shifting is fair use under § 107.") (citations omitted);

**Security**. The proposed amendment would also pose significant security risks. The current exemption was predicated on an understanding that security was a critical issue to be addressed from the beginning of development of a TDM corpus, not something that could be delegated to every other researcher that might someday decide it would be interesting to obtain a copy. For example, the Register's 2021 decision highlights that the proponents agreed:

> to include a requirement that researchers consult with their institution's IT office as they assembled the corpus and to specify that the security measures implemented would be designed "to prevent dissemination, downloading, and unauthorized access, and to limit access to the corpus of circumvented works only to other researchers affiliated with qualifying institutions for purposes of collaboration or the replication and verification of research findings."[34]

Although the proponents assert that, under their proposal, the institution at which an independent researcher works must comply with security standards as defined in the regulations, the proliferation of databases of unprotected motion pictures and literary works in relatively free circulation inevitably would mean that individuals further down the distribution chain with no connection to the researchers that constructed the database or their institution would have less knowledge of the requirements applicable to, and less motivation to protect, an asset they were simply given. It would be difficult, if not impossible, for copyright owners to identify downstream recipients. If the recipients could be identified, there is no guarantee that they would provide information to a copyright owner regarding the nature of their highest-level security measures, as the exemption now requires.[35] And it is not apparent what legal remedies copyright owners might have against downstream recipients for failure to implement required security measures.

As contemplated by 37 C.F.R. § 201.40(b)(4)(ii)(B), MPA, on behalf of its members, made a reasonable request for information about applicable security measures from some of the institutions that employ individuals who wrote letters in support of the proponents' proposal. Not only did some institutions *refuse* to share information about their security measures, but some responded aggressively. For example, Brandon Butler of the University of Virginia provided a letter included as Appendix F to the proponents' initial comments. That letter refers to "many projects" that were "enabled" by "the current exemption."[36] However, counsel for the University of Virginia responded to MPA as follows:

---

SECTION 1201 RULEMAKING: SIXTH TRIENNIAL PROCEEDING TO DETERMINE EXEMPTIONS TO THE PROHIBITION ON CIRCUMVENTION, RECOMMENDATION OF THE REGISTER OF COPYRIGHTS, 107-26 (Oct. 2015), https://cdn.loc.gov/copyright/1201/2015/registers-recommendation.pdf ("2015 Rec.").

[34] 2021 Rec. at 116.

[35] 37 C.F.R. § 201.40(b)(4)(ii)(B).

[36] AA 2023 Comment at App. F, p. 1.

> Setting to one side doubts that a trade association is entitled to make such a request, the University has no information to provide. Upon reasonable search and inquiry, I am aware of no such corpus at the University of Virginia.[37]

Similarly, Matthew Sag of Emory University provided a letter included as Appendix I to the proponents' initial comments. That letter contends that the "exemptions granted in 2021 have enabled some researchers to use digital methods to analyze e-books and DVDs without fear of liability under section 1201."[38] In response to an inquiry from MPA, counsel for Emory wrote:

> [T]he regulation cited in the letter applies to reasonable requests made by "a copyright owner whose work is contained in the corpus[.]" 37 CFR § 201.40(b)(5)(ii)(B). Ms. Temple's letter has not identified any work (1) for which the Motion Picture Association is the copyright owner; or (2) that is part of a corpus of works created by Emory for which circumvention has occurred. No further response to this correspondence is warranted.[39]

Of course, MPA did not identify itself as a copyright owner (because it was acting on behalf of its members who are copyright owners), and a copyright owner cannot possibly identify specific works in a corpus that a researcher has created without notification to the copyright owner.

These responses demonstrate glaring flaws in the security provisions of the current exemption. The situation would be much worse if the Librarian expanded the current exemption as proposed, because as databases of unprotected motion pictures and literary works circulated relatively freely among academic institutions, more and more copies would be in the hands of people who had no knowledge of what was in them, and it would be more and more difficult for copyright owners to get insight into how their valuable works were being exploited by people potentially far removed from creation of those databases.

**2. The proponents have not established that the additional activity that would be enabled by the broadened exemption would be a fair use.**

The proponents argue that all of this is just fine because, in their view, the Register's conclusion in the 2021 proceeding, along with the Second Circuit's decisions in *Authors Guild, Inc. v. HathiTrust*[40] and *Authors Guild v. Google, Inc.*[41] teach that TDM research is a "highly

---

[37] *See* Letter dated February 8, 2024 from Robert M. Tyler, Associate University Counsel, University of Virginia to Karyn Temple of the Motion Picture Association.

[38] AA 2023 Comment at App. I, p. 3.

[39] Email from Chris Kellner, Office of the General Counsel, Emory University to Katie Harper of the Motion Picture Association, dated February 12, 2024.

[40] 755 F.3d 87 (2d Cir. 2014).

[41] 804 F.3d 202 (2d Cir. 2015).

transformative" fair use.[42] However, their argument misses the point. A proper fair use analysis of the expanded exemption is fact-specific[43] and must take into account the full scope of activity that the expansion would enable. Neither the Register's 2021 decision concerning the carefully circumscribed current exemption, nor the *HathiTrust* or *Google* decisions addressing facts that involved little or no dissemination of the works involved, speak to the current proposal.

The Register's 2021 decision clearly stated that "the case law has *not* established that all copying of works for the purpose of TDM is necessarily a fair use."[44] Her analysis of the fair use issues was tightly bound to the details of the current exemption, including that "each researcher can utilize only the works in the corpus she has assembled" and "cannot aggregate her corpus with corpora assembled by other researchers."[45] She specifically highlighted that "decrypted copies can only be circulated to other institutions or researchers for the purpose of collaboration or replication and verification of research findings."[46] She also found "that the proposed exemptions demand close attention to security measures," noting that "[t]he courts that have found copying for the purpose of TDM to be fair use relied heavily in their analyses on the specific security measures that were in place."[47] In fact, she specifically conditioned her finding of fair use on TDM research "be[ing] conducted while maintaining the security of the copyrighted works in each corpus."[48] As described above, the proposed exemption (along with implementation of the security requirements in the current exemption) negates all of these fundamental underpinnings of the Register's 2021 decision.

The proponents' effort to link their proposal to *HathiTrust* and *Google* fares no better. Those decisions involved digitizing hard copy books owned by libraries without engaging in circumvention, acquiring unauthorized access, or engaging in widespread distribution of copies of the works involved to unrelated third parties. The activities in *HathiTrust* relevant to TDM involved full-text searching of a database of literary works, but display of only the page numbers and frequency of appearance of search terms—no distribution or display of text from the underlying work.[49] The defendant also took "extensive security measures" that played a major role in the court's analysis of market harm.[50] Similarly, *Google* involved providing search functionality that permitted users to identify books using search terms, view "snippets" of text containing search terms, and obtain statistical information.[51] The libraries whose books were scanned to include in Google's database received electronic copies of their books, but there was

[42] AA 2023 Comment at 19-23.

[43] *See, e.g., Harper & Row, Publishers, Inc. v. Nation Enterprises*, 471 U.S. 539, 549 (1985).

[44] 2021 Rec. at 107 (emphasis added).

[45] *Id*. at 108-09.

[46] *Id*. at 109.

[47] *Id* at 114.

[48] *Id* at 105.

[49] 755 F.3d at 91, 97. The court also addressed uses for the visually impaired and for preservation that are not relevant here.

[50] 755 F.3d at 100-01.

[51] 804 F.3d at 208-10.

no other distribution of the scanned works, and those libraries were required "to take precautions to prevent dissemination of their digital copies."[52] Even as it found that activity to constitute fair use, the court stated that it "test[ed] the boundaries of fair use."[53]

A different Second Circuit decision is more factually relevant here. *Fox News Network, LLC v. TVEyes, Inc.* involved a service that allowed users to search a text-searchable database of televised video programming, much like the databases of motion pictures that the proponents would like to enable researchers to distribute, and then to watch and download responsive video.[54] Focusing on the distribution of the works involved, the court recalled that *Google* "test[ed] the boundaries of fair use" and held that the defendant "exceeded those bounds."[55] So too here, the proposed broadening of the exemption to turn every researcher with an interest in TDM into a distributor of databases of unprotected motion pictures and literary works without effective security measures goes well beyond the activity that was found to be a fair use in *HathiTrust* and *Google* and exceeds the bounds of fair use.

### 3. This is neither the proceeding nor the time to break new legal ground in applying the fair use doctrine.

Recognizing that the proponents cannot rely upon the Register's 2021 conclusion concerning the TDM exemption or the *HathiTrust* or *Google* decisions as close analogies to what they propose here, this is neither the place nor the time for the Register to adopt a new interpretation of fair use capacious enough to cover the full scope of what the proponents propose. The Register has long held that proceedings like this are "not an appropriate venue for breaking new ground in fair use jurisprudence."[56]

Moreover, concerns applying to TDM have taken on new significance since the last rulemaking cycle due to rapid developments in the field of artificial intelligence ("AI"), which, like TDM, relies on the creation and use of large digital corpora that contain copyrighted works. In March of 2023, the Copyright Office launched an extensive initiative "to examine the copyright law and policy issues raised by [AI] technology" in which the issues raised by such corpora, or "datasets," used to train AI models figure prominently.[57] That initiative was the basis for four Copyright Office listening sessions that led in August 2023 to a Notice of Inquiry seeking comments on thirty-four separate questions related to AI and copyright, several with multiple distinct sub-parts, of which many concern such datasets.[58] Over 10,000 comments were submitted in that proceeding by interested parties addressing a wide range of copyright and

---

[52] 804 F.3d at 210-11.

[53] *Google*, 804 F.3d at 206.

[54] 883 F.3d 169, 173-75 (2d Cir. 2018).

[55] 883 F.3d at 174 (quoting *Google*, 804 F.3d at 206 (alteration in original)).

[56] 2021 Rec. at 10-11 (quoting Section 1201 Report at 117).

[57] *See* https://www.copyright.gov/ai/ (last visited Feb. 20, 2024); https://www.copyright.gov/policy/artificial-intelligence/ (last visited Feb. 20, 2024).

[58] Artificial Intelligence and Copyright: Notice of Inquiry and Request for Comments, 88 Fed. Reg. 59,942 (Aug. 30, 2023), https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf.

policy issues.[59] The ongoing consideration of such issues in a separate proceeding dictates extreme caution here.

The proponents have stated that "[t]he umbrella term 'TDM' is used internationally to refer to the use of copyrighted work in computational research,"[60] but have offered no useful guidelines to differentiate such "computational research" from areas presently under review in the Office's AI study. To the extent the proposed expansion would permit corpora to be distributed relatively freely among academic institutions to be used in the context of AI,[61] those issues are still under consideration in that separate proceeding. Given the Copyright Office's ongoing work in this area, an expanded Section 1201 exemption would be risky and premature.

* * *

As the Register has consistently reiterated, proponents "bear the burden of establishing that the requirements for granting the exemption have been satisfied."[62] In this instance, the proponents have completely failed to do so. The proposed expansion would thwart the carefully crafted current exemption, adopted after the proponents proffered the very language they attack here, and would expand access to vast databases of copyrighted works in that way greatly increasing the risk of copyright infringement at the very moment when the copyright implications of similar activity is under study.

Respectfully submitted:

/s/ J. Matthew Williams
J. Matthew Williams (mxw@msk.com)
Lucy Holmes Plovnick (lhp@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
1818 N Street, NW, 7th Floor
Washington, D.C. 20036
202-355-7904

*(signature block continues next page)*

[59] *See* https://www.regulations.gov/docket/COLC-2023-0006/comments (last visited Feb. 20, 2024).

[60] *See* Authors Alliance *et al.*, 2020 Long Comment at p. 4 (Dec. 14, 2020), https://www.copyright.gov/1201/2021/comments/Class%2007a%20and%2007b_InitialComments_Authors%20Alliance,%20American%20Association%20of%20University%20Professors,%20and%20Library%20Copyright%20Alliance.pdf.

[61] The letters in support of the proponents' proposed expansion make clear the strong relations between the fields of TDM and AI. *See* Long Comment at App. C, p. 1 (Letter of David Bamman, faculty member of the Berkeley Artificial Intelligence Research Lab (BAIR)); *id.* at App. J, p. 3 & n. 15 (Letter of Rachael Samberg and Timothy Vollmer, noting work with Professor Bamman "to obtain a grant to leverage TDM and artificial intelligence modeling," as some kinds of TDM research are "predicated upon machine learning for which artificial intelligence must first be trained").

[62] *See, e.g.*, 2021 Rec. at 7 (quoting 2015 Rec. at 13).

Robert H. Rotstein (rhr@msk.com)
James Berkley (jdb@msk.com)
Stacey Chuvaieva (stc@msk.com)
MITCHELL SILBERBERG & KNUPP LLP
2049 Century Park East, 18th Floor
Los Angeles, CA 90067
301-312-2000