



Charlesworth Law
15671 Royal Ridge Road | Sherman Oaks, CA 91403
jacqueline@charlesworthlaw.com | 917.432.7343

May 21, 2021

Via email: regans@copyright.gov; jrubel@copyright.gov

Ms. Regan A. Smith
General Counsel and Associate Register of Copyrights
Ms. Jordana S. Rubel
Assistant General Counsel
U.S. Copyright Office
101 Independence Avenue SE
Washington, DC 20559-6000

Re: Docket No. 2020-11
Exemptions to Prohibition Against Circumvention of Technological Measures Protecting
Copyrighted Works
Proposed Classes 7(a) (Motion Pictures – Text and Data Mining) and 7(b) (Literary
Works – Text and Data Mining)

Dear Ms. Smith and Ms. Rubel:

The Association of American Publishers (“AAP”) appreciates having had the opportunity to address Proposed Class 7(b) at the April 7, 2021 hearing before the Copyright Office. As you have observed, in reply comments filed prior to the hearing, petitioners modified their proposal to circumvent access controls on motion pictures and literary works for the purpose of engaging in text and data mining (“TDM”) activities. In addition, during the hearing, petitioners offered a number of further clarifications concerning the proposed exemption.¹ While AAP appreciates that the proposal to circumvent literary works is narrower than originally presented, as discussed below, significant issues remain. AAP therefore remains opposed to the adoption of Proposed Class 7(b), and its responses below should be understood as subject to its continuing objections, as set forth in its initial written comments and at the hearing.² *See generally* AAP’s Long Comment (Feb. 9, 2021) (“AAP Comments”).

AAP responds to the follow-up questions posed in your letter of April 16, 2021 as follows:

¹ As no transcript or video of the hearing is yet available, in responding to the Office’s questions, AAP is relying on notes taken during the hearing. In employing the term “petitioners” or “proponents,” AAP refers generally to supporters of the proposed exemption.

² AAP requests that the Office continue to consider its initial comments in relation to Proposed Class 7(b), including but not limited to the fair use concerns identified therein.



Charlesworth Law

15671 Royal Ridge Road | Sherman Oaks, CA 91403

jacqueline@charlesworthlaw.com | 917.432.7343

1. *Please provide your views regarding minimum, yet sufficient, security measures with which eligible institutions should be required to comply when creating a corpus of literary works or motion pictures on which text and data mining techniques can be performed. We welcome specific examples of standards for information security management currently used by academic institutions that the Office should consider, as well as suggestions of specific security measures that could potentially be used individually or in combination with other measures. We also invite you to compare standards you suggest with the approach taken by the EU Directive on Copyright in the Digital Single Market.*

As discussed in its initial comments and further below, AAP believes that Proposed Class 7(b) presents deeply troubling security risks on account of the amount and breadth of circumvention it would permit, as well as the resulting storage of unprotected works in uncontrolled, unsupervised environments. *See* AAP Comments at 3-4, 6-7. There is simply no precedent to support mass circumvention, reproduction and storage of a wide range of copyrighted works in the manner that petitioners propose.

Notably, the cases relied upon by petitioners, *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) (“*HathiTrust*”) and *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015) (“*Google Books*”), were careful to invoke the extensive security protocols employed by HathiTrust and Google in reaching their fact-limited determinations of fair use. *See* AAP Comments at 6-7. In AAP’s view, to the extent petitioners claim that their proposed TDM activities qualify as fair use under these precedents, the security measures to be applied to the resulting corpora of copyrighted works should be no less rigorous. Indeed, proponents’ own witness, Chris Hoffman of the University of California, Berkeley, testified to as much at the hearing, characterizing the contemplated collections of copyrighted works as “highly sensitive” material that would require multiple levels of physical and electronic security. He indicated that his university, like others, would look to standards published by the National Institute of Standards and Technology (“NIST”) to develop and enforce IT security guidelines in relation to such activities.

Mr. Hoffman provided examples of the security measures his institution would find necessary to protect the TDM research corpora being proposed by petitioners. These included: having the researcher sign an agreement with respect to the TDM project; registration of the project with the school’s information security office; deployment of human resources to assist with security; staff trainings on cybersecurity; physical security measures, including cameras and locks; encryption measures; system intrusion protections; and “choking” mechanisms to prevent reproduction of stored materials. Mr. Hoffman explained that there were 17 different security “control areas” to be considered. For an overview of the security protocols employed by Berkeley, *see* Berkeley Information Security Office, “Minimum Security Standards for Electronic Information,” <https://security.berkeley.edu/minimum-security-standards-electronic-information#baseline-profile-summary> (last visited May 20, 2021).



Charlesworth Law

15671 Royal Ridge Road | Sherman Oaks, CA 91403

jacqueline@charlesworthlaw.com | 917.432.7343

In light of the uncontested record, AAP believes that as a minimum baseline, any institution hosting the proposed TDM activities and resulting corpora must adhere to a security protocol that complies with NIST standards applicable to highly sensitive material. *See* NIST, “Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations,” *available at* <https://csrc.nist.gov/publications/detail/sp/800-171/rev-2/final> (last visited May 20, 2021). From this baseline, the hosting institution should also, of course, adopt additional safeguards as appropriate to particular projects and corpora. Among other protections, these must include mechanisms to detect and prevent downloading of stored materials, as well as destruction of the research corpora as soon as the TDM work is complete.

As is implicit in the above, AAP believes it would be entirely inappropriate and potentially disastrous to allow individual researchers to circumvent and create corpora of copyrighted works other than through and under the auspices of a NIST-compliant institution. Accordingly, should the Office decide to grant an exemption in this area, AAP believes the exemption should specify that any permitted TDM activities and hosting of content be conducted by and under the supervision of an institution of higher education that adheres, at a minimum, to NIST standards applicable to highly sensitive material. As noted, the security protocol must include automated protections to guard against the downloading of copyrighted material. Finally, as an additional, obvious security measure, any research corpus created under the exemption should be destroyed once the research project and any validation thereof are complete.

Not surprisingly, the European Directive on the Digital Single Market (“EU Directive”) repeatedly references the importance of security measures to protect copyrighted works being exploited for purposes of TDM. *See* EU Directive recital 15 (“[T]he copies should be stored in a secure environment.”), recital 16 (“[R]ightholders should be allowed to apply measures when there is a risk that the security and integrity of their systems or databases could be jeopardised.”), art. 3(2) (“Copies of works ... shall be stored with an appropriate level of security ...”), art. 3(3) (“Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted.”). At the same time, the EU Directive is a document designed to articulate overarching principles concerning the use of copyrighted works in the digital environment, rather than to provide specific language to effectuate those principles. The operative rules are left to the individual member states to legislate. *See* European Commission, “Applying EU Law,” at https://ec.europa.eu/info/law/law-making-process/applying-eu-law_en (directives are to “be incorporated by EU countries into their national legislation”) (last visited May 20, 2021). Except insofar as it emphasizes the need for appropriate security, AAP does not find the EU Directive to be especially helpful in establishing a specific, reliable standard for purposes of Proposed Class 7(b).



2. *Proponents explained in the hearing that the proposed text and data mining techniques would not enable researchers to view the text of literary works or the images from motion pictures included in a corpus in whole or in part. Please provide your views on regulatory language that would specify that researchers would not be permitted to view the text or images from works included in a corpus.*

As discussed below, the proponents of Proposed Classes 7(a) and 7(b) have refused to provide a definition of TDM for purposes of the exemption. *See* Petitioners’ Long Reply Comment (Mar. 10, 2021) (“Petr. Reply”) at 16 (“Petitioners are ... unwilling to artificially constrain eligible methods here.”) The lack of a clear definition for TDM is highly problematic for purposes of crafting a workable exemption to a statutory prohibition.

At the hearing, proponents represented that the TDM activities in which they seek to engage would be computational and nonconsumptive in nature and, as noted by the Office, would not require researchers to access text or images within a research corpus. Proponents also confirmed that the TDM output would not include expressive content from the circumvented works. As applied to literary works, AAP understands this to mean that (i) researchers would not be accessing any expressive content from the works contained in the research corpus, and (ii) the TDM output would merely identify specific instances and locations of the researcher’s search terms within the texts being analyzed, without reproducing textual content or images. AAP believes both of these limitations to be essential should the Copyright Office proceed to grant an exemption in this area. The ability to view or otherwise exploit the expressive content of circumvented works—including for the purpose of literary analysis—would substitute for paid copies of the works and thus be inconsistent with fair use. *See* AAP Comments at 8-9.

In this regard, we note that, according to petitioners’ current proposal, in order to circumvent a work for inclusion in a research corpus, the work must first have been “lawfully obtained” by the researcher or the researcher’s affiliated institution. *See* Petr. Reply at 7-9. Setting aside the problems with the “lawfully obtained” standard advocated by petitioners (discussed below), in the case of literary works, it seems petitioners contemplate that researchers would have independent access to full-text versions of the books they are circumventing—that is, they would not need to validate or analyze text within the research corpus itself. *See id.* at 22 (researchers “do not need th[e] exemption to read or watch works they already have lawfully obtained, and they do not need this exemption to quote from those works”). Thus, if the TDM research yielded a list of pages or locations on or at which a particular word or phrase appeared within a book, the researcher would consult a previously acquired version of the book to verify or evaluate those occurrences. Proponents’ recognition that researchers should already have independent, full-text access to the works they seek to circumvent underscores why any exemption can and should preclude access to expressive content through the research corpus.



Equally important, and consistent with petitioners' explanation at the hearing, there is no need for TDM results to display expressive content from the circumvented works, except for perhaps the titles of particular works. For literary works, the output should be limited to a list of pages or locations within each work where the researcher's search terms appear. *See, e.g.*, HathiTrust results pages for non-public domain "limited (search-only)" works, *available at* <https://www.hathitrust.org/> (last visited May 20, 2021). Indeed, this point was emphasized by proponents at the hearing, who explained that the output would consist of data *about* the works in the corpus, rather than content drawn from the works themselves.

3. *In their reply comments, proponents amended their proposed exemptions significantly in response to points opponents raised in their comments. These amendments introduced several new issues into the proceeding to which opponents have not have the opportunity to respond in writing. Opponents may respond to any new issues raised in proponents' reply comment.*

Although narrowed in certain respects, petitioners' amended requested exemption is still extraordinarily far-reaching. Petitioners' sweeping proposal—which would apply to every genre of literary work, apart from computer programs—is simply not supported by the record evidence. Moreover, the proposal fails to articulate meaningful safeguards to protect the copyrighted works that would be subject to circumvention. AAP remains extremely concerned about the prospect of an exemption that would permit individual researchers to circumvent thousands of copyrighted works—or more—for any activity the researcher deems "TDM."

Petitioners have failed to establish an adequate basis in fact or law to grant Proposed Class 7(b). A list of AAP's most serious concerns with the current proposal follows.

a. Refusal to define TDM

It bears repeating that, confronted with the concern that they have failed to identify the TDM activities in which they seek to engage, petitioners blithely assert that they are "unwilling to artificially constrain eligible methods here." *Petr. Reply* at 16. Petitioners continue to insist that the door be left open to an undefined list of TDM activities, current and future, consisting of unknown "diverse methods." *Id.* at 15.

AAP does not see how the Copyright Office can be expected to exempt an activity from the prohibition of section 1201 without a definition of that activity. Where the Office is tasked with articulating a legally binding regulation, it is inappropriate to decline to specify the activity that would be exempted. If the Office is, nonetheless, inclined to grant an exemption for TDM activities conducted on literary works, we urge the Office to adopt a definition that fits within the bounds of current U.S. law on fair use and specifically excludes access to content within the



research corpus, as well as research output that reproduces such content, as further discussed below.

b. Failure to limit proposal to literary works of fiction

Proponents initially sought to include every type of literary work in Proposed Class 7(b). As detailed in AAP’s initial comments, in addition to ebooks—the conceded focus of proponents’ submissions—petitioners’ extraordinarily broad proposal would encompass journals, periodicals, subscription and other databases, and website content and software, among other types of works classified as literary works under the Copyright Act. This kitchen-sink approach stands in direct contrast to the mandate that to qualify for an exemption, a proposed class must consist of a “narrow and focused subset” of works. *See* AAP Comments at 3. In their reply comments, petitioners abandoned their request to circumvent computer programs and “compilations ... compiled specifically for TDM purposes.” *Petr.* Reply at 6-7. These minor concessions do little to rein in an essentially boundless class. Petitioners’ proposal should be rejected for the simple reason that they have failed to propose a narrowly tailored class, as required under section 1201.

Perhaps not surprisingly, petitioners have failed to produce evidence to support their far-reaching proposal. The record with respect to Proposed Class 7(b) pertains only to ebooks and is silent as to everything else. Petitioners introduced no evidence to justify a need to circumvent journals, periodicals, databases or websites; insofar as they concern literary works, the letter exhibits attached to their petition focus on academic scholarship addressed to in-copyright works of fiction. Indeed, at the hearing, petitioners’ counsel Erik Stallman affirmed that the purpose of the proposal is to enable humanities scholars to analyze literary texts.

As explained in AAP’s initial comments, in order to qualify for an exemption, a proponent must be able to point to “distinct, verifiable, and measurable” adverse effects on the legitimate use of copyrighted works. AAP Comments at 10. The burden is on those who seek an exemption to demonstrate that the statutory criteria have been met. *See* U.S. Copyright Office, *Section 1201 Rulemaking: Seventh Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention* 12 (2018). These requirements must be established by a preponderance of the evidence before the Register can grant an exemption. *Id.* at 13.

Here, with respect to Proposed Class 7(b), the only impact claimed by petitioners involves asserted limitations on humanities scholarship in relation to works of fiction. With respect to other categories of literary works, the record is silent. *See* AAP Comments at 8 (“In what manner would researchers be engaging with periodicals, databases or software? And to what end? Petitioners offer no basis whatsoever to support their claim that these uses would be fair.”).



Accordingly, in keeping with section 1201 rulemaking standards, any exemption in response to Proposed Class 7(b) would need to be limited to TDM activities conducted on works of fiction in the form of books. There is simply no record evidence to support a broader exemption extending to nonfiction, journals, periodicals, databases or websites.

c. Failure to exclude subscription-accessed works and databases

A broad exemption is also in tension with petitioners' stipulation on reply that their proposal is meant to cover only works "lawfully obtained" by the researcher or the institution with which the researcher is affiliated. *Petrs. Reply* at 7. The term "lawfully obtained" is murky to say the least—based on petitioners' description, it seems it could refer to permitted access ranging from temporary viewing to a permanent copy. Indeed, petitioners acknowledge that "lawfully obtained" is intended to mean something broader than actual ownership. *Id.* at 8-9. In sum, "lawfully obtained" is a vague standard that does not definitively exclude circumvention of works accessed through a subscription or otherwise on a nonpermanent basis.

As discussed below, AAP submits that if the Office were to grant an exemption, either the researcher or affiliated institution through which the research is being conducted must *own* a copy of the work, as ownership is understood under relevant judicial precedent. *See, e.g., Krause v. Titleserv, Inc.*, 402 F.3d 119 (2d Cir. 2005) (setting forth test for license versus sale); *Vernor v. Autodesk, Inc.*, 621 F.3d 1102 (9th Cir. 2010) (same). In elaborating on the term "lawfully obtained," petitioners at one point suggest that it signifies that a user has paid a "one-time fee" for "permanent access" to the work. *See Petrs. Reply* at 8. AAP hopes this means petitioners do not in fact seek to circumvent temporary copies of works or copies that are accessed only through a subscription. Significantly, in this regard, we note that petitioners claim they are not proposing a "rent-to-own" model. *See id.* at 8.

It is confusing, then, that petitioners also assert that the proposed exemption should extend to databases—other than databases "compiled specifically for TDM purposes," whatever that means³, *see id.* at 7—because journal, scientific and other types of research databases are typically available only on a subscription basis. Proponents' position on databases is inconsistent with their claim that they are not seeking to generate permanent copies of works accessed through subscriptions.

Without question, any exemption should exclude subscription or temporarily accessed copies of copyrighted works. Among other concerns, if the work is accessed through a subscription service, either the researcher or the institution with which they are affiliated likely agreed to contractual terms that prohibit or restrict TDM activities. *See AAP Comments* at 10-11.

³ Would Westlaw or LEXIS fall into this category, for example? It is impossible to tell under the vague standards advocated by petitioners.



Regardless of whether an exemption exists, the researcher and/or the institution with which he or she is affiliated would still be bound by those licensing terms. In this regard, some subscription offerings include the ability to conduct TDM activities under the license, but for security purposes require the researcher to assemble and access the research corpus through the publisher’s designated API. *See id.* at 10. In other cases, the subscription is specifically for the purpose of facilitating TDM research of a collection of works. *See id.* For all of these reasons, subscription databases are particularly ill-suited for unauthorized TDM activities.

Last but certainly not least, again, there is no evidentiary record in this proceeding to support an exemption that would permit circumvention of databases consisting of journals, scientific research or other literary material. If the Office is considering a TDM exemption, it should not extend to databases.

d. Problematic use of “interfere”

Under proponents’ revised language, the exemption for TDM research would be available whenever an access control “interfere[s] with” TDM activities. *Petr. Reply* at 6. This is not an acceptable standard under section 1201. As discussed in AAP’s initial comments, an exemption is only appropriate where the TPM causes a material adverse impact on the claimed noninfringing use—a mere inconvenience does not qualify. *See AAP Comments* at 10. Accordingly, should the Office be inclined to adopt an exemption in this area, AAP submits that the regulatory language must be clear that circumvention is only permissible when the TPM is causing a material adverse impact on a noninfringing use.

e. Failure to limit beneficiaries of the exemption

Although petitioners’ revised proposal now specifies that a researcher seeking to avail itself of the exemption must be “affiliated with a nonprofit library, archive, museum or institution of higher learning,” *see Petr. Reply* at 6, 9-10, once again, their approach remains overbroad, imprecise, and unsupported by the record.

No evidence has been adduced in support of petitioners’ suggestion that non-university libraries, archives or museums should be permitted to engage in the proposed circumvention of literary works. Petitioners’ submissions and hearing testimony have focused solely on university communities. The Office should not extend a novel exemption presenting major security risks to innumerable museums, archives and non-research libraries where no record exists as to any adverse effects on these institutions or their patrons—or the capacity of such institutions to provide an appropriate level of security for substantial collections of circumvented copyrighted works.



Moreover, although addressed in proponents' reply comments, *see id.* at 10, the proposal itself does not specify that a researcher engaged in circumvention activities under the exemption must be a current faculty or staff member at the institution, or student working under that person's supervision. *See id.* at 6. This important qualification should be included in any exemption language, as to require only an "affiliation" would encompass nearly any association between an individual and an institution, no matter how attenuated or temporary. Any exemption should also expressly require that the institution of higher education with which the researcher is affiliated be the institution at which the circumvention and TDM activities will take place, and that is hosting and securing the circumvented content.

f. Failure to preclude all for-profit uses

Petitioners are now willing to stipulate that the immediate beneficiaries of the proposed exemption should be limited to nonprofit institutions and conducted "for the purpose of scholarly research and teaching." *Id.* at 6, 10-11. As observed in AAP's initial comments, even if research is initially, or primarily, conducted for scholarly or pedagogic purposes, it is not uncommon for university personnel to collaborate with, and provide research data or analysis to, for-profit enterprises. *See* AAP Comments at 8-9. The conduct or sharing of research for commercial purposes is at odds with a claim of fair use. *See id.* Accordingly, any exemption should specify that the TDM activities and results will be used *solely* for the purpose of nonprofit scholarly research or teaching activities.

g. Failure to ensure appropriate security for circumvented works

As discussed in AAP's initial comments, the *HathiTrust* and *Google Books* cases were expressly limited to their facts, which facts included the extensive security measures implemented by the two entities to prevent misuse or misappropriation of the copyrighted works they copied and were hosting. *See id.* at 4-6. These measures were expressly relied upon by the courts in their fair use analyses and are equally critical here. Indeed, the security risks posed by Proposed Class 7(b) seem considerably greater, as proponents are asking for the Copyright Office's blessing to allow individual researchers without knowledge of, or access to, sophisticated IT protocols to assemble vast libraries of unprotected copyrighted works. As noted in our earlier submission, this concern was highlighted in a published article by proponent Matthew Sag, who opined that it would be "reckless" to give a graduate student unsupervised access to a collection of copyrighted materials, as is being proposed here. *See id.* at 6 (quoting Sag).

At the hearing, Mr. Hoffman, the IT security expert from Berkeley, testified repeatedly that the research corpora of copyrighted works contemplated by proponents constitute "high risk" and "highly sensitive" materials from an IT security perspective. As noted above, he identified a host of measures his university would have to take to protect such a collection.

Despite Mr. Hoffman’s guidance, Mr. Sag’s blunt warning, and the reviewing courts’ approach in *HathiTrust* and *Google Books*, petitioners have failed to condition their proposal on the availability and deployment of specific and adequate security measures, ignoring the protocols identified by their own witness Mr. Hoffman. *See* Petrs. Reply at 6 (researchers need only employ unspecified “reasonable” security measures). Nor have they altered their proposal to ensure that only compliant institutions—rather than individual researchers—would be permitted to create and host the TDM corpora. *Id.* At the same time, it is abundantly clear that individual faculty members, students and staff would not be equipped to implement the sorts of IT security measures that experts like Mr. Hoffman say would be required to safeguard collections of highly sensitive circumvented works. *See id.* at 13 (acknowledging that security measures such as those employed by HathiTrust exceed the resources of an individual researcher).

As discussed above, any institution at which researchers would be engaging in activities under the proposed exemption should be NIST-compliant, with appropriate access controls and mechanisms in place to guard against misappropriation of copyrighted materials, and all research corpora should be destroyed as soon the TDM research and any validation process are concluded. The longer the corpus exists, the greater the opportunity for unauthorized access and reproduction.

In sum, with respect to security, if the Copyright Office proceeds with a TDM exemption, AAP believes it is imperative that (i) only institutions adhering to a baseline of NIST-level security standards, as well as anti-downloading safeguards and other appropriate IT protocols for highly sensitive materials—and not individuals—be allowed to carry out the circumvention activities and host circumvented works; (ii) any researcher seeking to engage in TDM activities under the exemption be currently affiliated with the hosting institution as a faculty member, staff member, or appropriately supervised student; and (iii) any resulting corpus of copyrighted works be permanently destroyed as soon as the research and any follow-up validation process are complete.

Thank you for your consideration of the above.

Sincerely,



Jacqueline C. Charlesworth

Counsel for Association of American Publishers