

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

ITEM A. COMMENTER INFORMATION

This comment is submitted on behalf of the Association of American Publishers (“AAP”). AAP represents the leading book, journal and educational publishers in the United States on matters of law and policy, advocating for outcomes that incentivize creative expression, professional content and innovative educational materials. AAP’s members depend first and foremost on a rational and effective copyright system.

Represented by:

Jacqueline C. Charlesworth
Jamie Frieden
Charlesworth Law
15671 Royal Ridge Road
Sherman Oaks, California 91403
917.432.7343

jacqueline@charlesworthlaw.com
jamie@charlesworthlaw.com

ITEM B. PROPOSED CLASS ADDRESSED

Proposed Class 3(b): Literary Works – Text and Data Mining

ITEM C. OVERVIEW

Petitioners Authors Alliance, American Association of University Professors and Library Copyright Alliance (collectively referred to herein as “petitioners” or “proponents”) propose to expand the current exemptions set forth in 37 C.F.R. § 201.40(b)(4) and (5) to circumvent technological protection measures (“TPMs”) on motion pictures and literary works to deploy text and data mining (“TDM”) techniques for purposes of scholarly research and teaching. Petitioners’ Long Comment (“Petr. Comment”) at 5. Petitioners have filed a single long-form comment addressed to both proposed classes, class 3(a) (motion pictures) and class 3(b) (literary works). *Id.* Although AAP’s submission is focused on literary works, many of the concerns raised herein are relevant to both of the proposed classes.

Petitioners seek to broaden the TDM exemption adopted in 2021 (“2021 Rule”) to allow “sharing” of corpora of works circumvented by a researcher at one institution (“circumventing

Privacy Act Advisory Statement: Required by the Privacy Act of 1974 (P.L. 93-579)

The authority for requesting this information is 17 U.S.C. §§ 1201(a)(1) and 705. Furnishing the requested information is voluntary. The principal use of the requested information is publication on the Copyright Office website and use by Copyright Office staff for purposes of the rulemaking proceeding conducted under 17 U.S.C. § 1201(a)(1). NOTE: No other advisory statement will be given in connection with this submission. Please keep this statement and refer to it if we communicate with you regarding this submission.

researcher or institution”) with researchers at other institutions (“noncircumventing or third-party” researchers and institutions). *Id.* By “sharing” they mean “copying and distributing.” *See id.* at 23 (expanded exemption would permit noncircumventing institutions to “receive” copies of circumvented corpora from circumventing institutions), 31 (same).

Petitioners portray their request as a “limited” modification of the existing exemption, but it is far from that. There is in fact no restriction on the number of third-party researchers or institutions with which a corpus assembled at one institution could be “shared” under their proposal. As drafted, the only constraint would be that the third-party researchers (though not the institutions with which they are affiliated) would need to comply with the requirements of the current exemption. *Id.* at 5-6.¹

Petitioners’ proposed amendment is facially flawed in another respect. Under the current regulation, the distribution of a circumvented corpus by one institution to another would violate the exemption, as the regulatory language clearly prohibits an institution from disseminating circumvented works. *See* 37 C.F.R. § 201.40(b)(5)(i)(D) (institution must use security measures to prevent dissemination or downloading). Even apart from proponents’ confusing construct, however, there is ample reason to question whether the recipient third-party institutions would, or even could, be compliant with the requirements of the exemption, as discussed below.

Petitioners’ purported rationale for an amendment that would permit exponential unauthorized distribution of vulnerable corpora of copyrighted works is that they do not understand the degree to which such corpora can be “shared” with researchers at other institutions under the existing exemption. According to petitioners, researchers are “stymied by the uncertainty” concerning this aspect of rule. *Petr.* Comment at 8-10. This is puzzling, for the exemption language concerning third-party access seems straightforward:

The institution [must] use[] security measures to prevent further dissemination or downloading of literary works in the corpus, and to limit access to only the persons identified in paragraph (b)(5)(i)(A) ... [affiliated students or IT staff] or to researchers ... affiliated with other institutions of higher education solely for purposes of collaboration or replication of the research.

37 C.F.R. § 201.40(b)(5)(i)(D). It is also puzzling because it was petitioner Authors Alliance that proposed the language to the Copyright Office for adoption in the last section 1201 rulemaking (“2021 Proceeding”).

AAP strongly opposes the unnecessary, high-risk distribution of decrypted corpora of books and other literary works under the auspices of section 1201 to an unlimited number of researchers and institutions. The breadth of the existing TDM exemption is already without precedent; it allows for the creation and maintenance of potentially massive corpora of in-copyright books and

¹ Although not actually required under the terms of the proposed amendment, petitioners assert in their comment that the third-party institutions would need to comply as well. *Petr.* Comment at 5.

films without transparency or oversight under security standards that appear to be aspirational at best. As we understand petitioners' proposal, circumventing institutions could, after assembling these vast collections, distribute them to third-party researchers and institutions, which recipients could, in turn, distribute them to still others. This seems well beyond anything that could possibly be the subject of an appropriate exemption under section 1201.

Apart from the alarming security issues presented by such a regime, the letters of support submitted by petitioners suggest that the corpora and/or results of TDM research could also be (and seemingly are being) used for their expressive content, including for the development and training of generative AI systems. The legality of exploiting copyrighted materials without permission to develop and populate AI systems is currently the subject of numerous lawsuits and is far from established as a fair use. There is no basis under current law for the Copyright Office to endorse such activities as a byproduct of the 1201 rulemaking process.

Against these very significant concerns, proponents have failed to offer any *bona fide* reason why third-party researchers seeking access to particular corpora cannot collaborate with researchers at circumventing institutions under the current exemption, thus avoiding further replication and storage of large collections of unprotected works. The record shows that collaboration is taking place. In short, there is no demonstrated need for the expansion petitioners seek.

Indeed, the "evidence" presented in the letters support attached to petitioners' proposal was apparently funded and generated at the direction of petitioner Authors Alliance *for the specific purpose* of inducing the Copyright Office to expand the TDM exemption to allow dissemination of decrypted corpora among academic institutions. In December 2022, petitioner Authors Alliance, along with six institutions (five of which—Bowdoin, Berkeley, Dartmouth, Stanford and Temple—submitted letters of support of the petition) *received funding totaling \$1 million* from the Mellon Foundation for the express purpose of "supporting demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act." Mellon Foundation, *Text and Data Mining*, <https://www.mellon.org/search/text%20and%20data%20mining> (last visited Feb. 18, 2024) (listing grants) (emphasis added); Authors Alliance, *Announcing the "Text and Data Mining: Demonstrating Fair Use" Project* (Dec. 22, 2022), <https://www.authorsalliance.org/2022/12/22/announcing-the-text-and-data-mining-demonstrating-fair-use-project%EF%BF%BC/> (Mellon-funded work will pursue issue of collaboration by "actually sharing copies of the corpus"); *see also* Exs. 4, 5 (copies of the foregoing). Given the hundreds of thousands of dollars bestowed upon institutions supporting the Authors Alliance and its co-petitioners, it is hardly surprising that every letter of support complains that the existing exemption does not go far enough and circumventing entities must be permitted to copy and distribute corpora of decrypted works. Needless to say, the fact that institutions were paid to produce evidence for this proceeding casts significant doubt on the credibility of these claims.

Still more, it appears that researchers and institutions currently engaged in TDM activities are not complying with critical security measures as required by the exemption. Under the 2021 Rule, absent an agreement with copyright owners of literary works (which to AAP's knowledge

does not exist), circumventing institutions are to employ the same security measures for circumvented works as they use to protect their own highly confidential information. *See* 37 C.F.R. § 201.40(b)(5)(ii)(B). In addition, they are obligated to provide information to copyright owners about those measures upon reasonable request. *Id.*; *see also* U.S. Copyright Office, *Section 1201 Rulemaking: Eighth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention* (Oct. 2021) (“2021 Recommendation”) at 117.

To better understand petitioners’ security practices under the existing exemption for purposes of evaluating the current proposal, AAP, acting on behalf of its copyright owner members—who distribute the vast majority of commercially published books in the United States, as well as millions of scientific, medical and engineering articles across thousands of journals annually—sent letters to the TDM researchers who provided statements of support for the expanded exemption. *See* Ex. 2 (AAP letter). In response to AAP’s letter, however, *not a single supporting researcher or institution provided information that described—or even identified—the specific security procedures in place to protect corpora of circumvented literary works.* *See generally* Exs. 1 (chart summarizing responses to AAP letter), 3 (copies of responses). Several recipients asserted that they were not required to respond to AAP’s query or did not respond at all. Exs. 1, 3. The refusal of circumventing institutions to provide meaningful information concerning the security measures applicable to their TDM activities strongly suggests that none in fact were implemented.

AAP submits that the current exemption should be repealed, or at least suspended, pending a finding by the Copyright Office based on convincing evidence that users of the exemption are abiding by its security requirements. In addition, assuming the exemption is continued, its language should be amended to ensure that copyright owners, either directly or through an authorized representative, are able to verify that circumventing institutions have implemented the requisite procedures. Because copyright owners have no insight into the specific works residing in researchers’ corpora, the rule should further require that circumventing researchers maintain records of the works they are circumventing, including identification of the relevant project and researcher and at least the author, title and publisher of each such work. These records, too, should be available to copyright owners and their representatives upon request.² Such transparency is crucial if the exemption is to continue in force.

Finally (and again assuming the exemption remains in place), AAP recommends modifying the current language of the exemption in additional ways to protect against the other risks identified herein, including the concern that TDM activities conducted under the exemption could extend to the training and development of generative AI systems. The modifications of the existing exemption proposed by AAP are described below.

² The language of the current rule suggests that a copyright owner should know if its works are contained in a corpus created under the exemption but does not provide a means for the owner to acquire that information. *See* 37 C.F.R. § 201.40(b)(5)(ii)(B).

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

1. Ebooks and Online Journals Are Heavily Targeted for Piracy

AAP's members, book, journal and educational publishers, use access controls—or TPMs—to protect against unauthorized access to and infringement of their copyrighted works. TPMs benefit copyright owners and users because they enable the development and deployment of electronic content delivery systems. The TPMs employed by AAP's members allow consumers to access and enjoy desired content at a time and place, and via a platform, of their choosing.

Unfortunately, bad actors circumvent access controls and eliminate the protections for ebooks and other literary works. Once hacked, an unprotected work can be uploaded to and shared freely on pirate sites without remuneration to the rightsholder. Sites trafficking in stolen ebooks and journal articles include online distribution hubs, cyberlockers and “shadow” libraries of pirated materials. *See, e.g.,* Greycoder, *A List of the World's Largest Shadow Libraries* (Aug. 12, 2022), <https://greycoder.com/a-list-of-the-largest-shadow-libraries/> (providing links to illegal sites).

Z-Library, for example, contains millions of pirated ebooks, and has had many of its domains shut down by the Department of Justice. *See* U.S. Attorney's Office, Eastern District of New York, *Two Russian Nationals Charged with Running Massive E-Book Piracy Website* (Nov. 16, 2022), <https://www.justice.gov/usao-edny/pr/two-russian-nationals-charged-running-massive-e-book-piracy-website>; *see also* Association of American Publishers, *Comment from Lui Simpson, Senior Vice President, Global Policy for Association of American Publishers (AAP), on the DOJ's Recent Action in the Z Library Case* (Nov. 17, 2022), <https://publishers.org/news/comment-from-lui-simpson-senior-vice-president-global-policy-for-the-association-of-american-publishers-aap-on-the-doj-s-recent-action-in-the-z-library-case/>. The notorious website Sci-Hub enables users illegally to access scholarly articles. According to the U.S. Trade Representative, Sci-Hub and its mirror sites facilitate unauthorized access “to over 88 million journal articles and academic papers, which comprise at least 90% of all toll-access published journal articles, a proportion greater than what is available legally to major institutional subscribers.” U.S. Trade Representative, *2023 Review of Notorious Markets for Counterfeiting and Piracy* 30 (2023), https://ustr.gov/sites/default/files/2023_Review_of_Notorious_Markets_for_Counterfeiting_and_Piracy_Notorious_Markets_List_final.pdf. In another example, pirate site Library Genesis (“Libgen”) claims to host copies of more than 80 million magazine articles, 2.4 million nonfiction books and 2.2 million fiction books, a vast number of which are infringing. *See* Library Genesis, *Libgen – Library Genesis* (Feb. 4, 2024), <https://libgen.onl/>.

What is more, AI platforms are being trained on corpora of pirated books without permission from or compensation to rightsholders. The dataset known as “Books3,” for example, which has been used by Meta and others to train AI models, comprises “pirated e-books spanning all genres.” Leah Asmelash, *These books are being used to train AI. No one told the authors*, CNN (Oct. 8, 2023), <https://www.cnn.com/2023/10/08/style/ai-books3-authors-nora-roberts-cec/index.html#>; *see also* Rights Alliance, *Rights Alliance removes illegal Books3 dataset used to train artificial intelligence* (Aug. 14, 2023), <https://rettighedsalliancen.com/rights-alliance-removes-the-illegal-books3-dataset-used-to-train-artificial-intelligence/> (host of Books3

forced to take down dataset after efforts by Danish Rights Alliance); Simon Sharwood, *Authors Guild sues OpenAI for using Game of Thrones and other novels to train Chat GPT: Class action alleges pirated novels were fed into binary brainbox*, The Register (Sept. 21, 2023), https://www.theregister.com/2023/09/21/authors_guild_openai_lawsuit.

2. Petitioners Have Provided No Information Concerning Their Circumvention Activities

In responding to Item D of the Long Comment form concerning the relevant TPMs and methods of circumvention, petitioners state only that “[t]he technological protection measures and methods of circumvention at issue for [their] proposal include those measures and methods applicable to motion pictures and literary works distributed electronically.” Petrs. Comment at 19. Petitioners’ response is circular and meaningless. According to the instructions in Office’s form, the Copyright Office seeks to understand “the nature and basic operation of the relevant technologies, as well as how they are disabled or bypassed.” U.S. Copyright Office, *Long Comment Regarding a Proposed Exemption Under 17 U.S.C. § 1201* at 1 (blank form). AAP and its members are entitled to this information as well.

AAP respectfully requests that the Office require proponents to explain the specific TPMs they are circumventing and the methods by which they are doing it. How and from what sources are researchers accessing electronic texts for circumvention? What types of TPMs are involved, and what are the methods of disabling or bypassing them? Who is performing the circumvention—individual researchers, IT staff members, or others? Petitioners’ lack of transparency concerning their existing and proposed TDM activities means that AAP is unable fully to evaluate and respond to their proposal.

ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGEMENT USES

3. Petitioners Have Failed to Establish Adverse Impact

A party seeking an exemption under section 1201 must be able to point to “distinct, verifiable and measurable impacts” in order to demonstrate that access controls are having an adverse effect on legitimate uses of copyrighted works. 2021 Recommendation at 12 (quoting H.R. Rep. No. 105-551, pt. 2, at 37 (1998) (“Commerce Comm. Report”)) (cleaned up). In addition, the petitioner must show that the prohibition on circumvention is causing the asserted adverse effects and preventing the proponents from making noninfringing uses without circumventing access controls. U.S. Copyright Office, *Section 1201 Rulemaking: Sixth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention* 83 (Oct. 2015).

An exemption cannot be granted on the basis of “*de minimis* impacts.” 2021 Rulemaking at 12 (quoting Commerce Comm. Report at 37) (cleaned up). In keeping with Congress’ intent, the Register of Copyrights has stressed that “mere inconveniences” caused by the prohibition do not satisfy the rulemaking standard. *Id.* (quoting Staff of H. Comm. on the Judiciary, 105th Cong., *Section-by-Section Analysis of H.R. 2281 as Passed by the United States House of Representatives on August 4, 1998*, at 6 (1998)) (cleaned up).

As referenced above, in December 2022 petitioner Authors Alliance secured funding from the Mellon Foundation to the tune of \$1 million dollars to generate TPM projects at various institutions for the specific purpose of creating a record upon which petitioners could assert in this proceeding that the current exemption is insufficient. That is, the projects were tactical undertakings that presumably would not have happened otherwise. *See* Ex. 4 (listing Mellon grants). Notably, the true purpose of the Mellon funding effort—namely, to generate a favorable record for this proceeding—is nowhere disclosed in petitioners’ submission, including in the attached letter of support from the Foundation itself. *See* Petrs. Comment App. M (Mellon Foundation letter).

Despite these very deliberate efforts, however, petitioners did not come close to establishing that researchers are suffering adverse impacts with respect to TDM research. In fact, the supporting statements attached to their submission demonstrate exactly the opposite. Each and every researcher that submitted a letter confirmed that he or she was able to circumvent works and perform research under the existing exemption. *See, e.g.*, Petrs. Comment App. H at 3 (supporting letter of Hoyt Long of University of Chicago (“Long Letter”)) (“[T]he exemption has been immensely valuable as it currently exists.”), App. C at 1 (supporting letter of David Bamman of University of California, Berkeley) (“This current exemption has allowed us to carry out substantial research ”); *see also generally id.* Apps. B-E, G-L (researcher letters of support).

At the same time—and perhaps unsurprisingly given the above-described funding operation—each and every letter also complains that the current exemption is too limited because it prohibits circumventing researchers from “sharing” the corpora they create with researchers at other institutions. Petitioners and their supporters claim that the existing provision allowing collaboration is ambiguous and that researchers are “stymied by the uncertainty surrounding what is and what is not allowed in the current exemption’s rules for corpora sharing.” Petrs. Comment at 8-10. Thus, they say, an amendment is required so researchers don’t have to worry about the rule but instead simply dispatch their decrypted corpora to researchers at other institutions. *Id.* at 5, 5-10.

AAP does not see how the provision permitting collaboration with outside researchers is ambiguous. Petitioners’ supposed confusion regarding the regulatory language addressed to collaboration is especially mystifying given that it was petitioner Authors Alliance itself that proposed the language to the Copyright Office, which adopted it. *See* 2021 Recommendation at 116 & n. 645; Authors Alliance Class 7 *Ex Parte* Letter to Copyright Office at 5 (Aug. 9, 2021) (describing its proposal as a “good option”). Consistent with the Author Alliance’s proposal, the 2021 Rule provides that the circumventing institution must employ effective security measures to prevent further dissemination or downloading of literary works in the circumvented corpus, but can provide access to researchers at other institutions “for purposes of collaboration or replication of the [TDM] research.” 37 C.F.R. § 201.40(b)(5)(i)(D). In other words, under the applicable provision, the circumventing institution *may not* copy or distribute the corpus but *may* allow an outside researcher who is collaborating with (or who seeks to replicate the research of) a researcher at the circumventing institution to access the corpus that is hosted by the circumventing institution.

Even if the provision were somehow ambiguous, as claimed, petitioners' fix amounts to burning down a house to kill a fly. A simple explanation from the Copyright Office of the requirements for and limits of collaboration among researchers would do.

Notwithstanding the litany of complaints, the letters of support in fact demonstrate that collaboration is not only possible, but happening. Researchers of motion pictures at Bowdoin and University of Rochester, for example, were able to collaborate on film analysis. *See generally* Petrs. Comment Apps. E (supporting letter of Joel Burgess and Emily Sherwood of University of Rochester), G (supporting letter of Allison Cooper of Bowdoin College). Notably, there is no submission from a researcher at a third-party institution complaining that he or she was unable to undertake or collaborate on a particular TDM project under current exemption, as one might expect to see if this were a legitimate concern.

In sum, the scenarios discussed in the supporters' letters are speculative and hypothetical, or at most suggest that it may require some logistical effort on the part of the circumventing institution to arrange for access to their corpora through a collaborative relationship as contemplated under the existing exemption.³ A nonverifiable or *de minimis* impact does not amount to an adverse impact unless it actually prevents a noninfringing use. Proponents' mere complaints, repetitious though they may be, do not amount to an adverse impact for purposes of section 1201.

4. The Copyright Office's Prior Determination of Fair Use No Longer Applies

In reaching her determination that the circumvention activities proposed by petitioners in the 2021 Proceeding could constitute a transformative fair use, the Register relied on two critical factual assumptions: (1) the resulting corpora would be protected by robust and effective security measures, and (2) the TDM uses would consist of algorithmic search processes and would not involve access to expressive content except for the very limited purpose of confirming research results. *See* 2021 Recommendation at 114 (question of fair use depends heavily on appropriate security), 109, 111 (content of works would be searched using algorithmic techniques, with any viewing of content solely to verify research). The current record indicates that users of the exemption are not abiding by these requirements.

a. Proponents' failure to implement security measures negates fair use

Under the existing exemption, researchers are able to circumvent virtually any electronically distributed literary work, including all in-copyright works of fiction and nonfiction from the 20th and 21st centuries. *See* 37 C.F.R. § 201.40(b)(5). As the Copyright Office acknowledged in the

³ Professor Hoyt Long at the University of Chicago asserts that his lab has been approached by third-party researchers seeking to access corpora at his institution, and that the requests have been turned down due to a lack of resources or subject matter expertise. He did not furnish any specifics or documentation to support these claims, however, or explain whether the requestors proceeded with their projects at their own or other institutions. Even if the claims were verifiable, they are no more than *de minimis* on this record and point to lack of expertise, resources and/or willingness to collaborate rather than an inability to circumvent TPMs. *See* Long Letter at 3.

2021 Proceeding, the corpora resulting from such circumvention activities could contain many thousands, or even millions, of works. *See* 2021 Recommendation at 114.

The ramifications of having third-party institutions create and host—and, if their current petition were to be granted, reproduce and distribute—enormous corpora of DRM-free literary works without transparency or oversight cannot be overstated. The 2021 Rule is explicitly premised on the implementation of appropriate security measures to safeguard circumvented works, as required under the fair use precedents relied upon by the Copyright Office in granting the exemption: *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*”). *See* 37 C.F.R. § 201.40(b)(5)(i)(D) (“The institution [must] use[] effective security measures to prevent further dissemination or downloading of literary works in the corpus, and to limit access....”).

The Register’s analysis of the security concern is worth quoting at length:

The Register concludes that the proposed exemptions demand close attention to security measures. The corpora envisioned by proponents could potentially contain hundreds, thousands, or even more copyrighted works. The 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. In *HathiTrust*, libraries implemented extensive security measures to reduce the risk of a data breach, including protecting the servers on which complete copies of the text of the copyrighted works were stored, and restricting and monitoring network access to the corpus. In *Google Books*, Google had walled off the copyrighted works from internet access and applied the same “impressive” security measures Google used to keep its own confidential information safe. Likewise, the European Directive on the Digital Single Market, emphasizes the importance of security measures to protect copyrighted works being used for TDM. Here, the potential for damage if there were to be a security breach is too great to allow researchers to implement security measures that they subjectively find to be reasonable without providing guidance as to at least minimum standards.

2021 Recommendation at 114-15 (footnotes omitted).⁴

The Librarian of Congress, as well, understood that security issues were paramount in approving the 2021 Rule:

⁴ In the case of *HathiTrust*, security procedures included “rigorous” physical security controls, “highly restricted access” to the corpus by library staff, “highly restricted” web access and protocols to prevent downloading of non-public domain works, and a “mass download prevention system” to shut off user access in case of excessive export activity. *HathiTrust*, 755 F.3d at 100 (citing Joint Appendix). As for *Google Books*, as noted by the Register, the book corpus was walled off from the internet and protected by the same “impressive” security measures used by Google to keep its own confidential information safe. *Google Books*, 804 F.3d at 228.

Most importantly, the Register recommended requiring the institution of higher education storing or hosting a corpus of copyrighted works to implement either security measures that have been agreed upon by copyright owners and institutions of higher education, or, in the absence of such measures, those measures that the institution uses to keep its own highly confidential information secure.

Final Rule, 86 Fed. Reg. 59627, 59633 (Oct. 28, 2021).

Given the overwhelming importance of these issues in relation to the current proceeding, in an effort to understand the security protocols currently in place, AAP sent a letter to each researcher who provided a statement of support for an expanded exemption who appeared to be engaged in circumvention activities, requesting a response by February 7, 2024. AAP's letter, attached hereto as Exhibit 2, asked for information concerning security measures implemented to protect literary works being circumvented under the current exemption, as permitted under 37 C.F.R. § 201.40(b)(5)(ii)(B). As provided in the 2021 Rule, the letter explained that the security measures used to protect circumvented corpora are supposed to be the same as those employed to protect the circumventing institution's own highly confidential information. Ex. 2; *see also* 37 C.F.R. § 201.40(b)(5)(ii)(B). The letter further explained that AAP was acting on behalf of publisher members who are copyright owners of literary works, a list of which owners was attached, and that the requesting parties represented the vast majority of the U.S. market for commercial books. In other words, any corpus of literary works created under the exemption would be extremely likely to contain works belonging to publisher members of AAP.

Of the ten institutions seeking an expanded exemption contacted by AAP, not a single one provided any information that confirmed implementation of safeguards to protect circumvented literary works. Five of these institutions (Dartmouth, Emory, University of Richmond, University of Rochester and Temple) declined to provide a substantive response. *See* Exs. 1, 3.⁵ Two (Berkeley and Chicago) offered up vague statements indicating that policies existed at their institutions but did not explain what policies, if any, were being applied to circumvention activities under the exemption. *Id.* Another two (Bowdoin and University of Virginia) asserted they were not aware of circumvention activities at their institutions involving literary works. *Id.* The final institution (Stanford) claimed to have no knowledge of the recipient researcher's circumvention of literary works under the exemption despite the professor's letter of support saying he was engaged in such activities. *Id.*; *see also* Petrs. Comment App. B at 1 (supporting letter of Mark Algee-Hewitt of Stanford University). In short, the responses yielded no reassurance whatsoever that circumventing institutions are taking the security requirements of the 2021 Rule seriously.

Based on this record, the existing exemption should be repealed, or at least suspended with respect to any institution that is unable to document that the institution (including affiliated researchers and staff) is in compliance with the security requirements of the exemption. And there is certainly no case for expansion of the exemption given proponents' utter failure to demonstrate that institutions are willing or able to comply.

⁵ Three of these said they needed more time beyond the February 7 deadline but never sent a response. *See* Exs. 1, 3.

Robust security safeguards were essential to the fair use determinations in *HathiTrust* and *Google Books*. An unwillingness or inability to abide by the security standards adopted by the Copyright Office as a condition of engaging in the circumvention activities at issue tips the scale decisively against fair use—with respect to the existing exemption and with respect to any proposed expansion of that exemption.

b. Proponents’ activities involve exploitation of expressive content, which is contrary to fair use

The 2021 Rule does not define “text and data mining.” *See* 37 C.F.R. § 201.40(b)(5). In the 2021 Rulemaking, proponents described TDM as an “umbrella term ... used internationally to refer to the use of copyrighted work[s] in computational research.” Petitioners’ Long Comment in 2021 Rulemaking Comment at 4 n.1. In the current rulemaking, petitioners do not define TDM, but attach a letter of support from Professor Matthew Sag, who defines it as follows: “Text data mining is an umbrella term referring to computational processes for applying structure to unstructured electronic texts and employing statistical methods to discover new information and reveal patterns in the processed data.” *Petr. Comment App. I* at 2.

In promulgating the 2021 Rule, the Copyright Office embraced petitioners’ definition of TDM activities as processes focused on computational analysis of the corpus of works, rather than analysis of expressive content. As the Office explained the TDM process: “The researcher would enter the algorithms she has designed and, using TDM techniques, the full content of each of the works within the corpus would be searched and the relevant information extracted in the form of a numerical score.” 2021 Recommendation at 109. The Office compared these activities to learning the frequency of usage of selected words in a text using Google’s ngrams tool or identifying the page number on which a particular appears through a HathiTrust search. *Id.* at 109-10 (internal quotes omitted).

The 2021 Rule thus provides that the content of works in the corpus cannot be viewed by circumventing researchers for purposes of analyzing that content but “solely for the purpose of verification of the research findings.” 37 C.F.R. § 201.40(b)(5)(i)(C). This limitation is to “ensure that the purpose of the use differs from the original expressive purposes for which the works were created.” 2021 Recommendation at 111. In this regard, it is worth emphasizing that reading and analysis of a textual work is an exploitation of expressive content of that work even if the acts are performed in an academic setting.

The record in this proceeding indicates the “TDM” activities of at least certain researchers go beyond narrow computational activities to develop statistical information about works and will involve exploitation of the expressive content of works. For example, the letter of support submitted by researchers at the University of Rochester indicates that motion pictures contained in their “Mediate” corpus are analyzed through a process called “close viewing,” which “involves human users manually annotating audiovisual materials.” *Petr. Comment App. E* at 1. In another example, at Bowdoin College’s “Kinolab,” student curators “closely watch” circumvented DVDs to annotate them and “pull clips” of scenes. *Id.* App. G at 2. The Bowdoin letter suggests that its corpus of circumvented films could serve “as an especially useful set of

training data ... [for] machine learning.” *Id.* App. G. at 3. A researcher at the University of Chicago is using circumvented literary works “to pursue experiments in the application of neural machine translation models to the analysis of newly digitalized collections of contemporary literature in several languages.” *Id.* App. H at 2. He observes that an expanded exemption would allow his current “digitized collections” to be used to develop and test large LLMs. *Id.*

Neither *HathiTrust* nor *Google Books* provides a sound basis for these sorts of activities. In neither case were full-text copyrighted works made available to users. Except in connection with the use of assistive technologies by print-disabled persons, HathiTrust did not display text from the copyrighted works in its corpus. *HathiTrust*, 755 F.3d at 91. As for Google, it limited the display of text in response to user searches to “snippet views” of about three lines of text; this limitation on users’ ability to view the book was critical to the court’s analysis. *See Google Books*, 804 F.3d at 210, 222-23, 226.

A subsequent fair use case decided by the Second Circuit, *Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169 (2d Cir. 2018), which involved a television news clipping service, underscored the distinction between a search function that identifies where a word appears in a news program and a service that allows users to view a clip in which the word appears. *See id.* at 176-77. The court of appeals held that the latter use, while perhaps “modest[ly]” transformative, was not a fair use because it allowed users to see and hear the plaintiff’s news content. *Id.* at 180-81.

A similar distinction applies here: for purposes of fair use, viewing a textual work to analyze its expressive content has very different implications than performing computational research to extract statistical information about the work. In granting the current exemption, the Copyright Office relied on petitioners’ representations that content would be viewed solely for the purpose of confirming computational research. *See* 2021 Recommendation at 110-11. It is therefore concerning—and inconsistent with fair use—that at least some researchers are accessing corpora to analyze, interpret and/or catalog expressive content, which uses are substitutional for the purpose of the original.

Even more concerning is the fact that researchers operating under the exemption are seeking to exploit their DRM-free corpora to train or develop generative AI systems. This is not a legitimate use of the exemption. But as acknowledged in some letters of support, circumvented corpora are useful in training and developing machine learning systems, including generative AI systems.

A significant amount of AI development has taken place in university settings. *See* Andy Baio, *AI Data Laundering: How Academic and Nonprofit Researchers Shield Tech Companies from Accountability*, Waxy (Sept. 30, 2022), <https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/> (“It’s become standard practice for technology companies working with AI to commercially use datasets and models collected and trained by non-commercial research entities like universities or non-profits”). For example, and highly relevant in the present context:

The academic researchers who compiled [a] Shutterstock dataset acknowledged the copyright implications in their paper, writing, “The use of data collected for

this study is authorised via the Intellectual Property Office’s Exceptions to Copyright for Non-Commercial Research and Private Study.”

But then Meta is using those academic non-commercial datasets to train a model, presumably for future commercial use in their products.

Id.

The exploitation of copyrighted works to train and develop generative AI systems constitutes expressive use of those works. As copyright and technology scholar Benjamin Sobel explains:

Emerging applications of machine learning challenge ... the[] premises of non-expressive use. First, machine learning gives computers the ability to derive valuable information from the way authors express ideas. Instead of merely deriving facts about a work, they may be able to glean value from a work’s expressive aspects; as a result, these uses of machine learning may no longer qualify as non-expressive in character.

Benjamin L. W. Sobel, *Artificial Intelligence’s Fair Use Crisis*, 41 Colum. J. L. & Arts 45, 57 (2017).

Since the 2021 Rule was adopted, numerous lawsuits have been filed that challenge the unlicensed use of copyrighted expressive content to build AI systems, including in academic and other non-profit settings.⁶ There is currently no precedent that allows this. Accordingly, there is no basis to rely on fair use to engage in generative AI activities under the TDM exemption.

c. The section 107 factors weigh against a finding of fair use

In the 2021 Proceeding, the Copyright Office decided that circumvention to conduct scholarly TDM activities could be a fair use of copyrighted works, subject to certain conditions. The record in this proceeding demonstrates that those conditions are not being met.

Section 107 of the Copyright Act prescribes the factors to be considered in evaluating a claim of fair use: (1) the purpose and character of the use, including whether it is of a commercial or nonprofit nature; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used; and (4) the effect of the use on the potential market for or value of the copyrighted work. 17 U.S.C. § 107.

In the prior proceeding, the Copyright Office concluded that engaging in the TDM activities proposed by petitioners for noncommercial, scholarly purposes could be a fair use *provided that any viewing or listening to copyrighted works was “solely to verify research results.”* 2021

⁶ To cite just a few: *Andersen v. Stability AI Ltd., et al.*, Case No. 23-cv-00201 (N.D. Cal) (complaint filed Jan. 13, 2023) (involving visual works); *Concord Music Group, Inc. v. Anthropic PBC*, Case No. 3:23-cv-01092 (M.D. Tenn.) (complaint filed Oct. 18, 2023) (song lyrics); *The New York Times Co. v. Microsoft Corp., et al.*, Case No. 1:23-cv-11195 (S.D.N.Y.) (complaint filed Dec. 27, 2023) (news content).

Recommendation at 108, 111. According to the Register, this limitation would “ensure that the purpose of the use differs from the original expressive purposes for which the works were created.” *Id.* at 111.

Under the Supreme Court’s decisions in *Andy Warhol Foundation for the Visual Arts, Inc. v. Goldsmith*, 143 S. Ct. 1258 (2023) (“*Warhol*”) and *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569 (1994), courts evaluating the first fair use factor are to consider whether the copier’s use is “transformative.” A transformative use is one that has a further purpose or different character, rather than serving as a substitute for the original. *See Warhol* 143 S. Ct. at 1276; *see also Campbell*, 510 U.S. at 579.

To the extent researchers are not limiting their TDM activities to algorithmic analyses and are accessing circumvented content to read it or analyze it directly, their use substitutes for the purpose of the original work and is nontransformative. The researchers described above who are engaged in “close watching” of film clips, translation of books, or development of machine learning models are engaging with expressive content and acting outside the boundaries of fair use as delineated by the Copyright Office in the 2021 Proceeding. The exploitation of expressive content for its intrinsic value tilts the first factor away from fair use.

The second fair use factor clearly favors copyright owners, since users are copying highly creative works at the core of copyright protection. The third factor also indisputably favors copyright owners, because researchers are copying whole works into their corpora.

The fourth factor of fair use, market harm, likewise weighs heavily in favor of copyright owners—even more so given legal developments since the last proceeding.

AAP members have developed and participate in licensing programs that permit researchers to carry out TDM activities. A leading example is the RightFind offering of the CCC, which makes millions of works available in a full-text format for TDM research by paying users. *See CCC, RightFind XML for Mining*, https://www.copyright.com/crc/wp-content/uploads/sites/2/RightFind-XML-for-Mining_FAQs.pdf (last visited Feb. 20, 2024). Since the last proceeding, demand for high-quality textual data has grown significantly due to the rise of generative AI and other machine learning models, which are often trained on books and other texts. Owners of large datasets are entering into agreements with companies seeking to use their data to build these systems. *See, e.g., Matt O’Brien, ChatGPT-maker OpenAI signs deal with AP to license news stories*, AP News (July 13, 2023), <https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a>; Universal Music Group, *Universal Music Group and Bandlab Technologies Announce First-of-its-kind Strategic AI Collaboration* (Oct. 18, 2023), <https://www.universalmusic.com/universal-music-group-and-bandlab-technologies-announce-first-of-its-kind-strategic-ai-collaboration/>. In short, the market for corpora of literary works that AAP described as nascent in the 2021 Proceeding is now rapidly developing.

As the Register explained in 2021, another aspect of the market harm analysis is “the nature and effectiveness of the security measures used to prevent unauthorized access to the corpus.” 2021 Recommendation at 107. The Register explained that the “potential for damage” in case of a security breach was “too great” to leave the matter of security up to individual researchers. *Id.* at

114-15. Accordingly, the Register imposed security requirements as a condition of exercising the exemption: in the absence of an agreement with copyright owners, circumventing institutions and researchers must protect circumvented literary works using the same security measures the institution uses to keep its own highly confidential information secure. 37 C.F.R. § 201.40(b)(5)(i)(D), (ii)(B). A failure to abide by this standard—as, based on the current record, appears to be the situation here—tilts the fourth factor decisively against fair use.

Petitioners claim that the recent *Warhol* decision supports their claim of fair use. In a remarkable assertion, petitioners proclaim “in *Warhol*, the Supreme Court affirmed that ‘[d]eriving uncopyrightable information and insights from copyrighted expression is not just transformative, it is highly transformative.’” *Petr. Comment* at 21. The *Warhol* decision does not contain the cited language, or anything like it. Petitioners instead misleadingly quote a statement made by Matthew Sag in his letter of support, *see id.* & n. 111; even the cited portion of Sag’s letter, however, references the *Warhol* decision. *See Petr. Comment App. I* at 2 & n.7 (supporting letter of Matthew Sag of Emory School of Law). In any event, because *Warhol* clarified and reined in the concept of transformative use, it does not help petitioners at all. *See generally Warhol*, 143 S. Ct. 1258.

Nor do developments on the international front lend support to petitioners’ mission. As the Register pointed out in her 2021 Recommendation, 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.⁷ 2021 Recommendation at 114-15. And notably, in 2023, the UK government abandoned a plan to adopt a broad copyright exception for TDM activities due to the harm it would cause to the creative sector. House of Lords, *Large language models and generative AI* ¶ 229 & n. 378 (Feb. 2, 2024), <https://publications.parliament.uk/pa/ld5804/ldselect/ldcomm/54/54.pdf>.

In sum, the Copyright Office has ample reason to reconsider its finding of fair use in the 2021 Proceeding in light of the evidence now before it. On the current record, the TDM activities engaged in by those operating under the exemption do not qualify as fair use.

5. Petitioners Have Failed to Meet Their Burden Under the Statutory Factors

Finally, petitioners have not satisfied the five statutory criteria to be considered for an exemption, as set forth in section 1201(a)(1)(C): (i) the availability for use of copyrighted works; (ii) the availability for use of works for nonprofit archival, preservation, and educational purposes; (iii) the impact that the prohibition on the circumvention of technological measures applied to copyrighted works has on criticism, comment, news reporting, teaching, scholarship,

⁷ *See also* 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.”), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0790>.

or research; (iv) the effect of circumvention of technological measures on the market for or value of copyrighted works; and (v) such other factors as the Librarian considers appropriate. 17 U.S.C. § 1201(a)(1)(C).

With respect to the first and second factors, the proposed exemption would discourage the dissemination and availability of copyrighted works in electronic formats by significantly increasing the probability that they would be exposed to piracy as a result of the proposed activities. The vast increase in copying and dissemination of decrypted corpora among researchers and institutions under petitioners' proposal without transparency or assurances that the corpora are properly protected compounds the risk exponentially. Copyright owners might choose to limit dissemination of electronic versions rather than take that risk.

Regarding the third factor, researchers already have sufficient means to collaborate with circumventing institutions to conduct scholarly TDM activities because it is already permitted under the existing exemption. Petitioners have offered no credible evidence of adverse impact on third-party researchers.

Concerning factor four, the widespread circumvention of literary works to build and disseminate unprotected collections of these works would devalue those works by undermining the legitimate market for the works, both in their original form and as included in licensed TDM datasets, which is inconsistent with fair use principles.

Finally, with respect to the fifth factor, AAP submits that the expanded exemption proposed by petitioners is at odds with the very purpose of section 1201, which is meant to encourage the digital dissemination of copyrighted works by allowing copyright owners to rely on access controls to protect their works. The exemption proposed by petitioners would severely undermine that goal.

6. Assuming It Remains in Place, the Current Exemption Should Be Modified to Protect Against the Concerns Described Herein

The record in this proceeding demonstrates that researchers circumventing literary works under the existing exemption are not adhering to the letter or spirit of the existing exemption. Should the exemption remain in place, AAP urges the Copyright Office to strengthen its language to emphasize its limitations and requirements so as to better protect copyright owners and their literary works, as set forth in the below redline.

37 C.F.R. § 201.40(b)(5)

(i) Literary works, excluding computer programs and compilations that were compiled specifically for text and data mining purposes, distributed electronically where:

(A) The circumvention is undertaken by a researcher affiliated with a nonprofit institution of higher education, or by a student or information technology staff member of the institution at the direction of such researcher, solely to deploy text and data mining techniques on a corpus of literary works for the purpose of scholarly research and teaching and no further purpose;

(B) The copy of each literary work is lawfully acquired and owned by the institution, or licensed to the institution without a time limitation on access;

(C) ~~Any~~ ~~The~~ person undertaking ~~the~~ circumvention or research activities views the contents of the literary works in the corpus solely for the purpose of verification of ~~the~~ statistical research findings and no other type of analysis; and

(D) The institution uses effective security measures to (i) prevent further dissemination or downloading of literary works in the corpus, and (ii) ~~to~~ limit access to only the persons identified in paragraph (b)(5)(i)(A) of this section or to researchers ~~or to researchers~~ affiliated with other institutions of higher education solely for purposes of collaboration or replication of the research where such access does not involve dissemination or downloading of works.

(ii) For purposes of paragraph (b)(5)(i) of this section:

(A) An institution of higher education is defined as one that:

(1) Admits regular students who have a certificate of graduation from a secondary school or the equivalent of such a certificate;

(2) Is legally authorized to provide a postsecondary education program;

(3) Awards a bachelor's degree or provides not less than a two-year program acceptable towards such a degree;

(4) Is a public or other nonprofit institution; and

(5) Is accredited by a nationally recognized accrediting agency or association.

(B) The term “effective security measures” means security measures that have been agreed to by interested copyright owners of literary works and institutions of higher education; or, in the absence of such measures, those measures that the institution uses to keep its own highly confidential information secure. If the institution uses the security measures it uses to protect its own highly confidential information, it must, upon a reasonable request from a copyright owner whose work is contained in the corpus or a representative of such owner, provide information to ~~that copyright owner~~ the requesting party regarding the nature of such measures, including specific information concerning the security measures applicable to activities conducted hereunder. In addition, any researcher or institution engaged in circumvention activities must maintain records of the literary works that are circumvented, including identification of the relevant project and researcher, and at least the author, title and publisher of each such work, and shall make such records available to copyright owners and their representatives upon reasonable request.

(C) The term “text and data mining” refers to computational research activities using algorithmic processes that do not involve the viewing of the contents of literary works other than as described in (b)(5)(i)(C) of this section and that result solely in statistical information about the literary works. Text and data mining does not include activities to train or develop machine learning systems or systems of artificial intelligence.

Conclusion

Proponents have refused to supply information to AAP’s publisher members regarding security measures they have adopted to protect the literary works they are circumventing, raising a strong inference they have failed to implement the safeguards required by the 2021 Rule. The failure to abide by prescribed security safeguards means that their TDM activities do not qualify for the existing exemption. But this has not stopped petitioners from seeking to enlarge the exemption in a dramatic fashion.

Even though the Copyright Office adopted petitioners’ proposed language to permit collaboration on TDM projects between researchers at different institutions, petitioners contend that the language is ambiguous and suggest that the only solution is to allow circumventing researchers to copy and disseminate potentially massive corpora of DRM-free works to other institutions, thus greatly magnifying the security risks inherent in the TDM exemption. Petitioners offer no *bona fide* reason why the circumventing institutions cannot enter into collaborative relationships with researchers at other institutions who seek to access their corpora, as contemplated by the current exemption. Moreover, their evidence of supposed adverse impact is suspect because petitioner Authors Alliance orchestrated a process whereby institutions received hundreds of thousands of dollars to support of the petition.

Section 1201 was not intended as a vehicle to have the Copyright Office sanction the creation, copying and dissemination of potentially massive libraries of decrypted literary works without permission of their owners or adequate security safeguards. The limited exemptions to the prohibition on circumvention are supposed to be narrowly crafted to avoid undue harm to copyright owners. Petitioners’ proposal goes well beyond any conceivable notion of fair use and the fair use claim would undoubtedly fail if litigated in court.⁸

⁸ “[T]here is no ‘rule of doubt’ favoring an exemption when it is unclear that a particular use is a fair or otherwise noninfringing use.” *Section 1201 Rulemaking: Seventh Triennial Rulemaking to Determine Exemptions to the Prohibition on Circumvention* at 15 (2018) (citing U.S. Copyright Office, *Section 1201 Rulemaking: Sixth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention* 15 (2015) and U.S. Copyright Office, *Section 1201 of Title 17* 115-16 (2017). In other words, “the rulemaking is not an appropriate venue for breaking new ground in fair use jurisprudence.” *Id.* (cleaned up).

DOCUMENTARY EVIDENCE

Please refer to Exhibits 1 through 5 hereto.

In addition, AAP requests that the online sources and information cited and/or linked to herein be considered in support of AAP's comment.

Dated: February 20, 2024

Association of American Publishers

By: /s/ Jacqueline C. Charlesworth

Charlesworth Law
15671 Royal Ridge Road
Sherman Oaks, CA 91403
Tel.: (917) 432-7343
jacqueline@charlesworthlaw.com

Counsel for Association of American Publishers

Exhibit 1:

Responses to AAP Letter Seeking Security Information on Behalf of Copyright Owners

Institution/ Addressee(s)/Respondent	Summary of Response
Bowdoin College Professor Allison Cooper	Says her lab does not engage in circumvention of literary works but does not address college as a whole
University of California at Berkeley Professor David Bamman, Ms. Rachael Samberg and Mr. Timothy Vollmer (Response from Principal Campus Counsel)	Objects that request did not come from copyright owners and that UC Berkeley is not required to provide information; provides a list of links to various UC policies without representing they are in use to protect circumvented works
University of Chicago Professor Hoyt Long	Claims to adhere to University of Chicago policies but says to follow up with Office of Legal Counsel for specifics of policies (no further response after letter forwarded to OLC)
Dartmouth College Professor John Bell	No response
Emory University School of Law Professor Matthew Sag (Response from university Associate General Counsel)	Objects that request came copyright owners through AAP rather than directly and does not identify specific works contained in an Emory corpus
University of Richmond Professor Lauren Tilton and Professor Taylor Arnold (Response from university General Counsel)	Want more time to respond
University of Rochester Professor Joel Burges and Mrs. Emily Sherwood, (Response from outside counsel Bond, Schoeneck & King)	Want more time to respond
Stanford University Professor Mark Algee-Hewitt (response from John E. Haugan, Senior University Counsel)	At present Office of General Counsel has no information that Algee-Hewitt is engaged in TDM activities involving post-19 th century works under the exemption (though Algee-Hewitt's letter in support of petition says he is)
Temple University Mr. Henry Alexander Wermer-Colan (Response from Associate University Counsel)	Want more time to respond
University of Virginia Mr. Brandon Butler (Response from Office of University Counsel)	Objects that request came from copyright owners through AAP rather than directly; claims UVA has no TDM corpus

Exhibit 2:
AAP Letter Seeking Security Information on Behalf of Copyright Owners



January 24, 2024

Re: Request for Information by Copyright Owners Pursuant to 37 C.F.R. § 201.40(b)(5)(ii)(B)

Dear Professor or Researcher:

The Association of American Publishers is the leading trade association representing the interests of book, journal and educational publishers in the United States in matters of law and policy. I write on behalf of our member publishers, a list of which is attached as Exhibit A (“Requesting Publishers”). Based on your recent statement submitted to the U.S. Copyright Office in support of expanding the current exemption permitting circumvention of technological measures on literary works to conduct text and data mining (“TDM”) activities, set forth in 37 C.F.R. § 201.40(b)(5) (“Exemption”), it appears that you and/or the institution with which you are affiliated are relying on the Exemption to engage in such activities.

The Requesting Publishers are the copyright owners and distributors of the majority of commercially distributed, copyright-protected literary works in the United States, which works are undoubtedly included in many of the TDM research corpuses created in reliance on the Exemption. It is our understanding that none of the Requesting Publishers has reached an agreement with you or your institution concerning the security measures to be applied to circumvented works contained in such corpuses. Accordingly, under the governing regulation, 37 C.F.R. § 201.40(b)(5), to qualify for the Exemption, the institution must employ the same security measures it uses “to protect its own highly confidential information.” 37 C.F.R. § 201.40(b)(5)(ii)(B). As permitted under the relevant provision, I therefore ask on behalf of the Requesting Publishers that you make arrangements with your institution to provide us with information concerning the nature of those measures, including:

1. An overview of the policies and protocols in place at your institution concerning the security measures that apply to your institution’s own highly confidential information (“HCI”).
2. Copies of any written policies or procedures concerning the security measures that apply to your institution’s HCI, including but not limited to:
 - a. The data security standard(s) in effect at your institution with respect to HCI
 - b. Written agreements with and/or registration of persons seeking or with access to HCI
 - c. Password, two-factor and/or other identification and authentication protocols applicable to persons with access to HCI
 - d. Screening, training and supervision of persons who have access to HCI
 - e. Protection of HCI storage facilities, including physical safeguards such as cameras and locks
 - f. Intrusion protections and alarms for systems that store HCI or through which HCI may be accessed
 - g. “Choking” or other mechanisms to prevent unauthorized reproduction of HCI
 - h. Monitoring procedures to identify and address security lapses that could threaten HCI

3. An explanation of the specific security measures applied by your institution to researchers and other persons seeking to engage in TDM activities under the Exemption.
4. Any written policies or procedures applicable to researchers or other persons seeking to engage in TDM activities under the Exemption.
5. An explanation of how applicable security measures are communicated to and enforced with respect to researchers and other persons engaged in TDM activities under the Exemption.
6. An explanation of any efforts to monitor and ensure compliance with applicable security measures by researchers and other persons engaged in TDM activities under the Exemption.
7. Any policies or procedures concerning the continued maintenance or destruction of TDM corpuses following completion of the relevant research.

If your institution has no information to provide in response to any one or more of the of the above queries, the response should so state. If your institution has *not* engaged in circumvention of any literary work of a Requesting Publisher under the Exemption, we ask that your institution provide a written statement confirming same.

We request a complete written response by **Wednesday, February 7, 2024**. The response should be addressed to me at thart@publishers.org. Thank you in advance for your cooperation.

Sincerely,



Terrence Hart
General Counsel
Association of American Publishers

Exhibit A

1517 Media

Academic Innovations

American Anthropological Association (AAA)

American Association for Cancer Research (AACR)

American Chemical Society (ACS)

American College of Cardiology

American Dental Association Publishing (ADA)

American Diabetes Association

American Institute of Aeronautics and Astronautics (AIAA)

American Medical Association (AMA)

American Physical Society (APS)

American Physiological Society, The (APS)

American Psychological Association (APA)

American Reading Company

American School for Classical Studies at Athens (ASCSA)

American Society of Anesthesiologists

American Society of Clinical Oncology (ASCO)

American Society of Health-System Pharmacists (ASHP)

American Society of Mechanical Engineers

American Thoracic Society (ATS)

American University in Cairo Press

Aquila Polonica (U.S.) LTD

Austin Macauley Publishers Inc.

Bar Charts, Inc.

Baylor University Press

Berghahn Books

BiggerPockets

Bloomsbury USA

Bookpress Publishing

Brookings Institution Press

Cambridge University Press

Candlewick Press

Center for Responsive Schools

Chronicle Books

City Lights Booksellers & Publishers

ClydeBank Media LLC

Collective Ink Limited

Columbia University Press

Common Ground Research Network

Davis Publications Inc.

Disney Publishing Worldwide

Edra Publishing LLC

eMATH Instruction Inc.

Emerald Publishing Group
F.A. Davis Company
Fordham University Press
Georgetown University Press
Getty Publications
Greystone Books Ltd
Group Publishing
Hachette Book Group
Hansen Publishing
HarperCollins Publishers Worldwide
Harvard Business Review Group
Harvard Education Publishing Group
Harvard University Press
Hogrefe Publishing
Holiday House Inc.
Humanix Books
Imbrifex Books
Independent Institute, The
INFORMS
Institute for Electrical and Electronics Engineers
International Monetary Fund (IMF)
John Wiley & Sons, Inc.
Johns Hopkins University Press
Judson Press
Kharis Publishing
Kodansha USA
Lynne Rienner Publishers, Inc
Macmillan
Massachusetts Medical Society
Materials Research Forum
McGraw Hill Education
MedMaster Inc
Melville House
Michigan State University Press
Moody Publishers
Morgan & Claypool Publishers
Museum of Modern Art , The (MOMA)
New World Library
New York University (NYU) Press
Okapi Educational Publishing, Inc.
Optica
Pearson
Penguin Random House
Pennsylvania State University Press
Princeton University Press
ProQuest, part of Clarivate

Quirk Books
RELX Group
Rosen Publishing Group
SAGE Publications
Scholastic, Inc.
Simon & Schuster
Society for Industrial and Applied Mathematics (SIAM)
Speedy Publishing
Springer Publishing Company
Stanford University Press
Tanglewood Publishing, Inc.
Taylor and Francis Group
Teacher Created Materials
Teachers College Press
Thames & Hudson
Turner Publishing Company
Universal Write Publications
University of California Press
University of Chicago Press
University of Massachusetts Press
University of Minnesota Press
University of North Carolina Press
University of Pennsylvania Press
University of Texas Press
University of Wisconsin Press
University Press of Colorado
Valancourt Books
W. W. Norton & Company
Walter de Gruyter, Inc.
Waterhouse Press
Wesleyan University Press
White House Historical Association
Yale University Press
YBK Publishers

Exhibit 3: **Responses to AAP Letter**

Bowdoin College

From: Allison Cooper <acooper@bowdoin.edu>
Sent: Wednesday, February 7, 2024 6:08:14 PM
To: Terrence Hart <thart@publishers.org>
Cc: Allison Cooper <acooper@bowdoin.edu>; Megan Hart <mhart@bowdoin.edu>
Subject: Re: Request for security information relating to TDM circumvention activities

Dear Mr. Hart,

I am writing in response to your letter of January 24, 2024. Kinolab, the digital humanities laboratory for the analysis of film language that I direct at Bowdoin College, does not engage in the circumvention of TPM on literary works to conduct text and data mining activities. As a faculty member, I cannot speak on behalf of the College as a whole, so please direct any further inquiries you may have to Megan Hart in Bowdoin's Office of Legal Counsel (copied above).

Cordially,

Allison Cooper (she/her/hers)
Associate Professor of Romance Languages and Cinema Studies
Bowdoin College
Brunswick, Maine 04011
tel: 207.798.4188; email: acooper@bowdoin.edu

University of California at Berkeley

From: Liv K Hassett <livhassett@berkeley.edu>
Sent: Wednesday, February 14, 2024 6:09:13 PM
To: Terrence Hart <thart@publishers.org>
Subject: Request for UCB Security Standards

Dear Mr. Hart,

I am Principal Campus Counsel for the University of California, Berkeley ("UCB"). I am in receipt of your letters dated January 24, 2024 that you addressed to a UCB faculty member (David Bamman) and two UCB staff members (Timothy Vollmer and Rachael Samberg) inquiring about UCB's security standards and policies. In the future, please direct all communications regarding this matter to me (or to the Public Records Coordinator, as noted below).

Your letter seeks documents and information purportedly under 37 C.F.R. § 201.40(b)(5) (the "Regulation"). That Regulation creates an information request right for copyright owners whose works are contained in research corpora that have been decrypted pursuant to the Regulation. You have not provided us with any information by which to evaluate whether you are a copyright owner whose works are contained in such corpora at our campus.

In addition, that Regulation limits a copyright owner's information request right to the "nature" of "the security measures [the institution of higher education] uses to protect its own highly confidential information." Your letter asks for a number of documents that exceed the scope of this limited request right.

Although you have not yet demonstrated that you are eligible to make requests under the Regulation, as a courtesy, please see the links below for the information contemplated by the Regulation.

- University of California's systemwide information security policy (<https://security.berkeley.edu/IS-3-informational-page>);
- UCB Data Classification Guideline (<https://security.berkeley.edu/data-classification-guideline>);
- UCB Secure Research Data and Computing (<https://research-it.berkeley.edu/services-projects/secure-research-data-computing>);
- UCB Guide to Working with Sensitive Data (<https://docs-research-it.berkeley.edu/services/high-performance-computing/user-guide/sensitive-data/>);
- UCB Guide to Classifying Research Data (<https://security.berkeley.edu/resources/how-classify-research-data>);
- UCB's Information Security Office researcher website (<https://security.berkeley.edu/resources/researcher-resources>).

If you wish to request further records from UC Berkeley, please contact UC Berkeley's Public Records Coordinator: pra@berkeley.edu.

Sincerely,

Liv Hassett

LIV K. HASSETT

she/her

Principal Campus Counsel

UC Legal - UC Berkeley Office of Legal Affairs

200 California Hall #1500

Berkeley, CA 94720-1500

(510) 926-9578 (mobile)

livhassett@berkeley.edu

/LEE/v

University of Chicago

From: Hoyt Long <hoytlong@uchicago.edu>
Sent: Monday, February 5, 2024 10:26 AM
To: Terrence Hart <thart@publishers.org>
Subject: Re: Request for security information relating to TDM circumvention activities

Dear Mr. Hart,

Good morning. I've reviewed your letter and request from a few weeks ago. As a researcher at the University of Chicago, I adhere to my institution's policies and guidelines on text and data mining activities.

If you would like to learn more about these policies, especially as they relate to data use and privacy, I encourage you to reach out directly to our [Office of Legal Counsel](#). They will be better able to respond to your request.

Best regards,

Hoyt

--

Hoyt Long
Professor of Japanese Literature
Chair, East Asian Languages and Civilizations
University of Chicago

Dartmouth College

[No response]

Emory University School of Law

From: Kellner, Chris <ckellne@emory.edu>
Sent: Friday, February 9, 2024 3:11:53 PM
To: Terrence Hart <thart@publishers.org>
Subject: Request for security information relating to TDM circumvention activities

Mr. Hart – the regulation you've cited applies to reasonable requests made by "a copyright owner whose work is contained in the corpus[.]" 37 CFR § 201.40(b)(5)(i)(B). Your letter has not identified any work (1) for which the American Association of Publishers 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 your letter is warranted.

Sincerely,

Chris Kellner

Christopher J. Kellner
Senior Intellectual Property Specialist and Associate General Counsel
Office of the General Counsel, Emory University
101 Administration Building, 201 Dowman Drive
Atlanta, GA 30322 | Tel (404) 727-6012
chris.kellner@emory.edu

From: Sag, Matthew James <matthew.james.sag@emory.edu>
Sent: Wednesday, January 24, 2024 10:55 AM
To: Terrence Hart <thart@publishers.org>
Subject: Re: [External] Request for security information relating to TDM circumvention activities

I have not read the attached letter. This seems like an inquiry properly directed to a university general Counsel.

<https://www.ogc.emory.edu/about/about-homepage.html>

Matthew Sag
Professor of Law Emory University

University of Richmond

From: Sinclair, Shannon <ssinclair@richmond.edu>
Sent: Friday, January 26, 2024 1:08 PM
To: Terrence Hart <thart@publishers.org>
Cc: Tilton, Lauren <ltilton@richmond.edu>; Arnold, Taylor <tarnold2@richmond.edu>
Subject: Request for security information relating to TDM circumvention activities

Dear Mr. Hart,

I serve as general counsel for the University of Richmond. Professor Tilton has forwarded your letter of January 24, 2024 to me. I ask that you direct all future correspondence regarding the University of Richmond to my attention. I will review your letter and will respond once I have completed my review.

Best regards,

Shannon E. Sinclair
Vice President and General Counsel
University of Richmond
110 UR Drive
Maryland Hall Suite 200
University of Richmond, Virginia 23173
(804) 287-6683 (o)
(703) 405-5164 (c)
ssinclair@richmond.edu

University of Rochester



One Lincoln Center | Syracuse, NY 13202-1355 | bsk.com

GEORGE R. MCGUIRE
gmcguire@bsk.com
P: 3152188515
F: 3152188100
C: 3152785604

February 7, 2024

VIA ELECTRONIC MAIL

Terrence Hart, Esq.
General Counsel
Association of American Publishers
1730 Pennsylvania Avenue NW
Washington, DC 20006

Re: *Request to University of Rochester for Information by Copyright Owners Pursuant to 37 C.F.R. § 201.40(b)(5)(ii)(B)*

Dear Mr. Hart:

We are counsel to the University of Rochester ("U of R"). Your letter to the U of R dated January 24, 2024, has been provided to us for review and response.

U of R is highly respectful of the intellectual property rights of others, including the copyrights held by the members of the Association of American Publishers, and certainly would not intentionally act in a manner that is in derogation of such rights. It therefore takes the request made in your letter very seriously and is investigating the matter to provide you with a response. As your letter requested a reply by February 7, 2024, we did want to advise of U of R's efforts to investigate and provide a formal response. However, as I am sure you can appreciate, investigating such a matter at a large research university such as the U of R takes time. We will therefore diligently investigate the matter and respond to your request in as timely a manner as possible.

In the meantime, should you have any questions, please do not hesitate to contact me with any questions or concerns.

Very truly yours,

BOND, SCHOENECK & KING, PLLC

George R. McGuire
Member

17234267.1 2/7/2024

Attorneys At Law | A Professional Limited Liability Company

Stanford University

STANFORD
UNIVERSITY

OFFICE OF THE GENERAL COUNSEL
John E. Haugen
Senior University Counsel



Telephone (650) 725-7575
Facsimile (650) 723-4323
haugenje@stanford.edu

February 19, 2024

Via Email

Dear Mr. Hart,

I am an attorney in the Office of General Counsel at Stanford University. Your letter dated January 14, 2024 to Professor Mark Algee-Hewitt was passed on to me.

I am looking into the situation and gathering information related to your request for information. At this time, the Office of General Counsel has no information suggesting Professor Hewitt, a researcher focused on 18th and 19th century literature, has engaged in circumvention of any literary work of a Requesting Publisher under the TDM Exemption. The January 14th letter includes no reference to specific activities by Stanford University or Professor Algee-Hewitt that would make a detailed response to your letter necessary or appropriate under the circumstances.

As an academic institution, we take the protection of intellectual property rights seriously. Should we be missing information available to you, or should we learn of anything additional, we would be happy to engage in a more detailed manner.

Sincerely,

A handwritten signature in black ink that reads "John E. Haugen".

John E. Haugen

Temple University

From: Jackie Rodriguez <jackie.rodriguez@temple.edu>
Sent: Monday, February 5, 2024 5:07:57 PM
To: Terrence Hart <thart@publishers.org>
Cc: Alex Wermer-Colan <alex.wermer-colan@temple.edu>
Subject: RE: [External] Request for security information relating to TDM circumvention activities

Dear Mr. Hart:

I am writing to advise that Temple is gathering information to respond to your January 24th letter. While we understand that you requested a response by February 7th, our office must coordinate with several other offices to respond to your inquiry and require additional time to do so. We appreciate your patience.

Best regards,
Jackie

Jackie Rodriguez (*she/her*)
Associate University Counsel
Temple University | Office of University Counsel
Cell: (267) 935-4969

University of Virginia



OFFICE OF THE
UNIVERSITY COUNSEL

ROBERT M. TYLER
ASSOCIATE UNIVERSITY COUNSEL
Direct: 434.297.7624
Email: robtyler@virginia.edu

February 8, 2024

By Email: thart@publishers.org

Terrence Hart
General Counsel
Association of American Publishers
1730 Pennsylvania Avenue NW
Washington, DC 20006

Re: AAP Request for Information Under 37 C.F.R. § 201.40(b)(5)(ii)(B)

Mr Hart:

Your January 24 letter to Brandon Butler cites his comments to the Copyright Office concerning the perceived chilling effect of copyright law on TDM activities. In his comments, which he expressly made in his personal capacity, Mr. Butler discussed the findings of an international survey of TDM researchers. I'm therefore at something of a loss why his comments would cause AAP to draw any conclusion about activities at the University of Virginia, let alone to demand that the University provide information about its purported reliance upon 37 C.F.R. § 201.40(b)(5). Thankfully, AAP's stated commitment to scholarship, free expression, and the democratic exchange of ideas allows me to conclude it wasn't an effort to chill either commentary or wholly legal TDM activity.

An institution is required to provide certain information in response to a reasonable request from a copyright owner whose work is contained in a corpus of literary works to which an institutional researcher has undertaken circumvention for scholarly TDM research and teaching. 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.

Any further communications to the University on this matter should be sent to me. Please reach out if you have any questions or would like to discuss.

Sincerely,

Robert M. Tyler

cc: Brandon Butler

Madison Hall | 1827 University Avenue | PO Box 400225 | Charlottesville, VA 22904-0225
P 434.924.3586 | F 434.982.3070 | www.virginia.edu

From: Terrence Hart <thart@publishers.org>
Sent: Friday, January 26, 2024 10:45 AM
To: Butler, Brandon (bcb4y) <bcb4y@virginia.edu>
Cc: Iler, Cliff (cgg8qy) <Cliff.iler@virginia.edu>; universityofvirginia@mycusthelp.net
Subject: FW: Freedom of Information Act Request :: V003483-012424

Dear Mr. Butler:

I received the below communication from the University of Virginia ("UVA") FOIA office, apparently in response to my recent letter on behalf of book publisher members of the Association of American Publishers ("AAP").

AAP's letter is not a FOIA request that merely seeks historical records. Rather, we are requesting specific answers to questions regarding UVA's protection of highly confidential information, as we are entitled to do under 37 C.F.R. § 201.40(b)(5). That provision sets forth the requirements of the text and data mining ("TDM") exception to section 1201's prohibition on anticircumvention activities. UVA is required to supply such information as a condition of engaging in such activities.

In any event, please note that the regulation in question, promulgated under title 17 of the United States Code, preempts Virginia state law with respect to activities involving the reproduction of copyright-protected literary works. See 17 U.S.C. §§ 102, 301(a). Further, even if it did apply (which it does not), the Virginia FOIA statute does not mandate the provision of the type of security-related information required under 37 C.F.R. § 201.40(b)(5). See Va. Code § 2.2-3705.2. In other words, the apparent referral to your FOIA office was inappropriate.

I hope this helps to clarify AAP's inquiry and that we will receive the requested information shortly. A failure to provide such information will mean that UVA is not in compliance with the requirements of the TDM exemption.

Sincerely,

Terry Hart
General Counsel
Association of American Publishers
C: 202-322-3858

From: University of Virginia FOIA <universityofvirginia@mycusthelp.net>
Sent: Wednesday, January 24, 2024 5:22 PM
To: Terrence Hart <thart@publishers.org>
Subject: Freedom of Information Act Request :: V003483-012424

--- Please respond above this line ---

Dear Terrence Hart:

I am writing in response to your recent request for records under the Virginia Freedom of Information Act (FOIA), Va. Code § 2.2-3700 et seq., for the following records:

- *1. An overview of the policies and protocols in place at your institution concerning the security measures that apply to your institution's own highly confidential information ("HCI").*
- 2. Copies of any written policies or procedures concerning the security measures that apply to your institution's HCI, including but not limited to:*
 - a. The data security standard(s) in effect at your institution with respect to HCI*
 - b. Written agreements with and/or registration of persons seeking or with access to HCI*
 - c. Password, two-factor and/or other identification and authentication protocols applicable to persons with access to HCI*
 - d. Screening, training and supervision of persons who have access to HCI*
 - e. Protection of HCI storage facilities, including physical safeguards such as cameras and locks*
 - f. Intrusion protections and alarms for systems that store HCI or through which HCI may be accessed*
 - g. "Choking" or other mechanisms to prevent unauthorized reproduction of HCI*
 - h. Monitoring procedures to identify and address security lapses that could threaten HCI*
- 3. An explanation of the specific security measures applied by your institution to researchers and other persons seeking to engage in TDM activities under the Exemption.*
- 4. Any written policies or procedures applicable to researchers or other persons seeking to engage in TDM activities under the Exemption.*
- 5. An explanation of how applicable security measures are communicated to and enforced with respect to researchers and other persons engaged in TDM activities under the Exemption.*
- 6. An explanation of any efforts to monitor and ensure compliance with applicable security measures by researchers and other persons engaged in TDM activities under the Exemption.*
- 7. Any policies or procedures concerning the continued maintenance or destruction of TDM corpora following completion of the relevant research.*

Under Va. Code §2.2-3704(A), "all public records shall be open to citizens of the Commonwealth, representatives of newspapers and magazines with circulation in the Commonwealth, and representatives of radio and television stations broadcasting in or into the Commonwealth during the regular office hours of the custodian of such records." If you do not meet either of these qualifications, your request may not be fulfilled.

If you are a member of the media, can you identify the newspaper or magazine with circulation in the Commonwealth that you represent in connection with this request, or can you identify the radio and television station broadcasting in or into the Commonwealth that you represent in connection with this request? If you are not a member of the media, pursuant to Virginia's FOIA statute, Va. Code §2.2-3704(A), the University of Virginia requires requesters to provide their full name and legal address. Please send me your name and legal address so that I may continue to process your request.

Sincerely,

Freedom of Information Act Officer

University of Virginia
www.virginia.edu/foia

Exhibit 4:
Mellon Foundation Grants to Fund Proponents of TDM Exemption

Source: Mellon Foundation website:
<https://www.mellon.org/search/text%20and%20data%20mining>

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee
Authors Alliance, Inc.
Location
Berkeley, California, United States
Amount
\$100,000
Date of award
December 08, 2022
Length
12 months
Grantmaking area
Public Knowledge
Guiding strategy
Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee
Bowdoin College
Location
Brunswick, Maine, United States
Amount
\$100,000
Date of award
December 08, 2022
Length
15 months
Grantmaking area
Public Knowledge
Guiding strategy
Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee

University of California at Berkeley

Location

Berkeley, California, United States

Amount

\$200,000

Date of award

December 08, 2022

Length

12 months

Grantmaking area

Public Knowledge

Guiding strategy

Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee

Dartmouth College

Location

Hanover, New Hampshire, United States

Amount

\$200,000

Date of award

December 08, 2022

Length

12 months

Grantmaking area

Public Knowledge

Guiding strategy

Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee

Stanford University

Location

Stanford, California, United States

Amount

\$100,000

Date of award

December 09, 2022

Length

18 months

Grantmaking area

Public Knowledge

Guiding strategy

Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee

Temple University

Location

Philadelphia, Pennsylvania, United States

Amount

\$200,000

Date of award

December 08, 2022

Length

12 months

Grantmaking area

Public Knowledge

Guiding strategy

Networks and Sharing

Text and Data Mining: Demonstrating Fair Use

to support demonstration projects to advance the argument that text and data mining of content constitute applications of fair use under the Digital Millennium Copyright Act

Grantee

University of Cincinnati

Location

Cincinnati, Ohio, United States

Amount

\$100,000

Date of award

December 09, 2022

Length

12 months

Grantmaking area

Public Knowledge

Guiding strategy

Networks and Sharing

Exhibit 5:
Announcement of Authors Alliance TDM Fair Use Project With Mellon Foundation

Source: Authors Alliance website:
<https://www.authorsalliance.org/2022/12/22/announcing-the-text-and-data-mining-demonstrating-fair-use-project%E2%9C%93/>

Announcing the “Text and Data Mining: Demonstrating Fair Use” Project

Posted December 22, 2022

We’re very pleased to announce a new project for 2023, “Text and Data Mining: Demonstrating Fair Use,” which is generously supported by the Mellon Foundation. The project will focus on lowering and overcoming legal barriers for researchers who seek to exercise their fair use rights, specifically within the context of text data mining (“TDM”) research under current regulatory exemptions.

Fair use is one of the primary legal doctrines that allow researchers to copy, transform, and analyze modern creative works—almost all of which are protected by copyright—for research, educational, and scholarly purposes. Unfortunately, in practice, not everyone is able to use this powerful right. Researchers today face the challenge that fair use is often overridden by a complex web of copyright-adjacent laws. One major culprit is Section 1201 of the Digital Millennium Copyright Act (“DMCA”), which imposes significant liability for users of copyrighted works who circumvent technical protection measures (e.g., content scramble for DVDs), unless those users comply with a series of specific exemptions to Section 1201. These exemptions are lengthy and complex, as is the process to petition for their adoption or renewal, which recurs every three years.

Text data mining is a prime example of work that demonstrates the power of fair use, as it allows researchers to discover and share new insights about how modern language and culture reflect on important issues ranging from our understanding of science to how we think about gender, race, and national identity. Authors Alliance has worked extensively on supporting TDM work in the past, including by [successfully petitioning the Copyright Office for a DMCA exemption](#) to allow researchers to break digital locks on films and literary works distributed electronically for TDM research purposes, and this project builds on those previous efforts.

The *Text Data Mining: Demonstrating Fair Use* project has two goals in 2023:

- 1) To help a broader and more diverse group of researchers understand their fair use rights and their rights under the existing TDM exemption through one-on-one consultations, creating educational materials, and hosting workshops and other trainings; and
- 2) To collect and document examples of how researchers are using the current TDM exemption, with the aim of illustrating how the TDM exemption can be applied and highlighting its limitations so that policymakers can improve it in the future.

We’ll be working closely with TDM researchers across the United States, as well organizations such as the Association for Computers and the Humanities, and will be actively exploring opportunities to work with others. If you have an interest in this project, we would love to hear from you!

About The Andrew W. Mellon Foundation

The Andrew W. Mellon Foundation is the nation's largest supporter of the arts and humanities. Since 1969, the Foundation has been guided by its core belief that the humanities and arts are essential to human understanding. The Foundation believes that the arts and humanities are where we express our complex humanity, and that everyone deserves the beauty, transcendence, and freedom that can be found there. Through our grants, we seek to build just communities enriched by meaning and empowered by critical thinking, where ideas and imagination can thrive. Learn more at mellon.org.