VIA EMAIL

May 21, 2020

Regan Smith, Esq.
General Counsel
U.S. Copyright Office
Library of Congress
101 Independence Avenue SW
Washington, D.C.20540

RE: Post-Hearing Response to Class 7(b)

Dear Ms. Smith:

Thank you for the opportunity to address the points made in the petitioners' request for text and data mining exemptions. SIIA's comments are confined to class 7(b). Petitioners initially proposed the following class:

> Class 7(b): *Lawfully accessed literary works distributed electronically where circumvention is undertaken in order to deploy text and data mining techniques.*

In our view, no response or opposition is needed to an exemption written in this fashion, as the Copyright Office has already made clear that such a proposal is invalid on its face as a "class of works."[1] Moreover, while text and data mining has been found to be fair use in very narrow circumstances,[2] these cases do not remotely suggest that the intonation of "text and data mining" leads automatically to a finding of fair use.[3]

---

[1] *See* U.S. Copyright Office, 2018 Register's Recommendation, at 13-14; H. Rep. No. 105-551, at 38 (requiring a class of works to be a narrow and focused subset of one of the statutory categories).

[2] *E.g.*, Author's Guild v. Hathi trust, 755 F.3d 87 (2d Cir. 2014); Author's Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

[3] *E.g., Google*, 804 F.3d at 222 (noting the importance of "blacklisting" certain kinds of works and portions thereof, and that the defendant had excluded certain works such as short poems, cookbooks,

Not only did the 2018 Register's Report and the DMCA's legislative history provide notice that such a proposal is legally unsound, but we also note that many of petitioners' additional refinements find ready analogues in the case law cited in their hundred-page filing. Based on the comments that counsel made at the hearing, this appears to have been a strategic choice that leaves their "real" proposal as the last word.[4] While these post-hearing letters and the ex parte process partially ameliorate the effect of that approach, we respectfully suggest that it may be more efficient to incentivize petitioners in this proceeding to craft facially valid proposals that provide both the Office and the public with reasonable notice of the relief petitioners seek.

In response to concerns raised in the opposition comments,[5] petitioners narrowed their proposal to read as follows:

> *Proposed Class 7(b) (revised): Literary works, excluding computer programs, distributed electronically and lawfully obtained, that are protected by technological measures that interfere with text and data mining, where: (1) the circumvention is undertaken by a researcher affiliated with a nonprofit library, archive, museum, or institution of higher education to deploy text and data mining techniques for the purpose of scholarly research and teaching; and (2) the researcher uses reasonable security measures to limit access to the corpus of circumvented works only to other researchers affiliated with qualifying institutions for purposes of collaboration or the replication and verification of research findings.*

This proposal appears to have been narrowed still further based on comments made during the hearing that the Office held in April.

Fair use and "text and data mining" is a rapidly developing state of the law around a series of practices that lacks a clear legal definition. SIIA urges the office to heed its own admonition that

---

and dictionaries); *Hathitrust,* 755 F.3d at 91 (noting that no text was disclosed to the user in snippet form or otherwise). *Compare*

[4] As a transcript of the hearing is not yet available, counsel is relying on memory and contemporaneous notes of what transpired.

[5] *See, e.g.*, Comments of the Association of American Publishers on Class 7(b).

there is no rule of doubt for issuing an exemption: if the case for the use is not clear, it should be denied.[6]

These statements raise a separate problem with the petition: causation. The statute's prohibition has nothing whatsoever to do with the difficulties that the petitioners have in accessing Hathitrust. Nor is it the cause of their perceived difficulty. SIIA believes that this was confirmed at the hearing, as petitioners stated that they lack the resources to do the scanning at the scale they would like and that access to Hathitrust facilities was cumbersome.[7] The works that petitioners seem to be concerned with were not "born digital" and with technological protection measures attached. They already exist in a suitable form for text and data mining, and the statute is not preventing either the petitioners' access to or copying of such works.

SIIA therefore continues to believe that the proposed exemption is unsupported by the record even in its narrower form and that is should not be issued for both legal and precedential reasons. In the event that the Register intends to recommend an exemption revolving around petitioners' second proposal, that proposal remains troublingly overbroad. The balance of this filing addresses that overbreadth.

I. The Class Should Include Only Those Works for Which Petitioners Have Demonstrated Interest

Although the petitioners narrowed the scope of the proposed class to exclude computer programs from literary works, the proposed class still applies to a number of SIIA-member products, including copyright-protected databases, newsletters, and scientific, technical and medical publications. Petitioners introduced no evidence whatsoever that section 1201 is interfering with fair use of these materials.

In addition, SIIA's members actually make their content available for computational analysis in different formats, depending on the user's needs. Academic users, for example, receive XML content so

---

6 2018 Register's Recommendation, at 15.

7 AAP Comments, at 11. See also, e.g., Authors Alliance Opening Comments, at 11-12 (noting the difficulty of access to Hathitrust). These concerns have nothing whatsoever to do with section 1201's prohibition.

that computational analysis can more easily be performed. They ask, however, that the researcher delete the XML versions once the analysis has been performed. In addition, particularly large (usually private sector) users may receive bulk feeds via secure FTP protocols, which is easier to keep current. Our members also apply rate limits to these users, will stop excessive use, and require them to delete the corpus once the work is finished. Similarly, SIIA has many educational, newsletter and other specialized publishers—many of whom sell access online—who would be affected by this proposal. Acknowledging the overbreadth of their initial ask, petitioners focused on a narrower ask: electronic books, specifically American fiction dated post-1945. If the Office is persuaded that an exemption is needed, then the Register should recommend an exemption only for those works for which it finds specific record evidence.

II. The "Lawfully Obtained" Language Should be Changed to Exempt Licensed Copies of Literary Works

The exemption must be expressly limited those who possess a copy of the work in which the copyright owner's distribution rights have been exhausted. The breach of a license condition is infringement, while the breach of a covenant is remedied by contract.[8] In both *Google* and *Hathitrust,* the defendants used works in which first sale had extinguished the right of distribution. The existing law of fair use does not permit a user to breach a license condition to text and data mine, and the copyright office would be making new law if it crafted an exemption that permitted it.

III. Text and Data Mining Must be Appropriately Defined

The term "text and data mining" is insufficiently precise and runs the risk of being used for any activity (including consumption) that a petitioner might claim. Thus, we would suggest that for the purposes of the exemption, the term of "text and data mining" should refer to the full-text searching of a covered work for the sole

---

[8] *See generally* 5 Patry on Copyright § 17:43; Sun Microsystems, Inc. v. Microsoft Corp., 188 F.3d 1115, 1122 (9th Cir. 1999).

purpose of computational or quantitative analysis of that work, provided that such analysis results in no part of the text of that work being perceptible by the researcher.

To the extent that text and data mining has been held to be fair use, that claim is strongest when the defendant does not permit access to the underlying corpus and allows only full-text search and quantitative analysis of the works contained within it.[9] During the hearing, petitioners and their counsel seemed primarily to be bemoaning the difficulties and inconvenience of performing this type of research using the Hathitrust servers, and would like to replicate that access in a more convenient form. They acknowledged that many instances of text and data mining do not require access to the actual text of these works, but only to the quantitative results of search queries. The definition should be narrowly confined to that activity.

IV. The Proposal Must Further Limit the Scope of Authorized Users of the Exemption and Require the Deletion of the Corpus Within Three Years of the Circumvention.

Petitioners have crafted their exemption to apply to "acts of circumvention" made by "researchers affiliated" with nonprofit library, archive, museum, or institution of higher education. As written, this language has three significant problems. The first is that the "affiliated" language is far too broad. On one side, for-profit institutions often formally affiliate with universities for any number of reasons. On the other, an alumnus can colorably claim to be "affiliated" with a university. And in either case, in the absence of an employee relationship, the institution may not be legally responsible for the acts of an "affiliated" individual—a problem compounded by the fact that the proposal allows the researcher to share the results with anyone affiliated with a qualifying institution. The proposal should therefore be limited to employees of a nonprofit institution performing the research within the scope of their employment.

---

[9] Cf. *Hathitrust*, 755 F.3d at 91 (reporting the number of times a particular word appeared in a scanned work, but not the underlying text).

The second problem with the language has to do with the duration of the statute of limitations, which is three years.[10] After that time, the limitations on use that attach to the circumvention are arguably inapplicable, and the institution's only constraint is infringement liability: it has possession of a huge corpus of works that it may use *beyond* the exemption and will not be subject to the limits contained within it. In addition, many of these institutions may well be state entities immune from suit under the Eleventh Amendment,[11] meaning that they have no responsibility to compensate copyright owners for the harm they may cause. We therefore suggest the exemption only be available to those entities that do *not* have Eleventh Amendment immunity from suit, and that in all cases the corpus must be deleted within three years of the date of circumvention.

Third, "scholarly research and teaching" must be the *sole* purpose for which this activity occurs. SIIA members are in litigation about the use of artificial intelligence to "mine" copyright-protected works in order to create competing products. Competitive activity is well beyond the scope of accepted fair use in existing law, and the exemption should categorically exclude it.

V. "Reasonable Security Measures" Must be Carefully Defined, Along with the Conditions of Access to the Works

Petitioner's revised proposal requires "reasonable security measures." First, part and parcel of these measures ought to be a requirement that the institutions contact the copyright owners or organizations that represent them about their plans to circumvent and seek their approval. Given the likely size of these undertakings and the nascent state of the law in this area, the copyright owners should be aware of who has stripped access protection from their works. We note that there is some precedent for this approach in the European Union.

Second, we believe that any exemption ought to expressly define such measures, and the hearing record should be helpful in listing factors that at a minimum ought to be met. Indeed, we note that both NIST has information security standards that ought to be

---

10 17 USC 507.

11 *See* Allen v. Cooper, 140 S. Ct. 994 (2020).

applicable.[12] Petitioners' witnesses noted that these requirements scale and made the (somewhat discomfiting) observation that smaller institutions often do a better job with cybersecurity than larger ones. They should therefore not be burdensome.

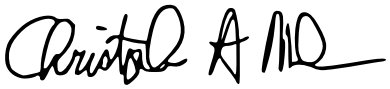
And finally, the full text of the works should not be accessible once they have been scanned and indexed for search. This limitation is consistent with the claimed purpose of the exemption, namely, the computational analysis of particular literary usages or trends in the works, and not the potentially substitutive reproduction of the works themselves.

Conclusion

As mentioned above, we do not believe that issuance of a text and data mining exemption is warranted. If the Register does decide to recommend one, we would hope that those who opposed it have at least an informal opportunity to offer feedback on it. And certainly, SIIA would welcome the opportunity to answer any questions that the Office may have about our views.

Thank you again for the opportunity to provide these additional views.

Sincerely,



Christopher A. Mohr
Vice President for Intellectual
Property and General Counsel

---

[12] See, e.g., National Institute of Standards and Technology, Protecting Controlled Unclassified Information in Nonfederal Systems and Organizations, https://csrc.nist.gov/publications/detail/sp/800-171/rev-2/final (providing standards for nonclassified information held by federal contractors).