



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

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

## ITEM A. COMMENTER INFORMATION

This Comment has been submitted on behalf of the Authors Alliance, the American Association of University Professors, and the Library Copyright Alliance.

(1) **Authors Alliance** is a nonprofit organization with the mission to advance the interests of authors who want to serve the public good by sharing their creations broadly. We create resources to help authors understand and enjoy their rights and promote policies that make knowledge and culture available and discoverable. For more information, visit <https://www.authorsalliance.org>.

*Represented by:*

Samuelson Law, Technology & Public Policy Clinic

UC Berkeley, School of Law

Erik Stallman and Catherine Crump, Faculty

Ziyad Alghamdi, Tait Anderson, and Erin Moore, Clinical Law Students

[estallman@clinical.law.berkeley.edu](mailto:estallman@clinical.law.berkeley.edu)

(2) **The American Association of University Professors (AAUP)** is a nonprofit membership association of over 45,000 faculty, librarians, graduate students, and other academic professionals. Since our founding in 1915, the AAUP has helped shape American higher education by developing the standards and procedures that maintain quality in education and academic freedom in this country's colleges and universities. We define fundamental professional values and standards for higher education, advance the rights of academics, particularly as those rights pertain to academic freedom and shared governance, and promote the interests of higher education teaching and research.

Risa Lieberwitz, AAUP General Counsel, [rlieberwitz@aaup.org](mailto:rlieberwitz@aaup.org)

Aaron Nisenson, AAUP Senior Counsel, [anisenson@aaup.org](mailto:anisenson@aaup.org)

Nancy Long, AAUP Associate Counsel, [nlong@aaup.org](mailto:nlong@aaup.org)

(3) **The Library Copyright Alliance (LCA)** consists of three major library associations—the American Library Association (ALA), the Association of College and Research Libraries (ACRL), and the Association of Research Libraries (ARL)—that collectively represent over 100,000 libraries in the United States.

**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 Web site 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.

*Represented by:*

Jonathan Band  
policybandwidth  
jband@policybandwidth.com

## TABLE OF CONTENTS

ITEM A. COMMENTER INFORMATION .....	1
ITEM B. PROPOSED CLASS ADDRESSED.....	3
ITEM C. OVERVIEW .....	3
ITEM D. TECHNOLOGICAL PROTECTION MEASURES AND METHODS OF CIRCUMVENTION .....	4
ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGEMENTG USES .....	5
1. The original exemption request is fully supported by law, but petitioners would not object to a narrower exemption provided that the core purposes of the exemption are preserved. ....	5
a. A more narrowly tailored exemption can address the bulk of opponents' concerns while still enabling the exemption's core purpose. ....	6
b. Other objections to the scope of petitioners' exemption request are without merit.....	15
2. TDM for the purpose of scholarly research and teaching is a fair use. ....	16
a. <i>HathiTrust</i> and <i>Google Books</i> are settled law and support the conclusion that the secondary use enabled by the exemption is fair use.....	16
b. TDM for scholarly research and teaching is a noninfringing fair use under the four statutory factors. ....	19
3. Section 1201 adversely affects noninfringing TDM research and there is no practical alternative to circumvention that eliminates that harm. ....	25
b. Commercial databases are too siloed and limited to be meaningful alternatives to circumvention. ....	26
d. Anti-circumvention liability is the cause of the adverse effect proponents seek to mitigate in this proceeding. ....	29

## **ITEM B. PROPOSED CLASS ADDRESSED**

This Reply comment addresses Class 7(a) Motion Pictures and Class 7(b) Literary Works.

## **ITEM C. OVERVIEW**

Petitioners seek an exemption for research using text and data mining (“**TDM**”), groundbreaking techniques and methods of computational analysis that allow researchers to ask questions about collections of literary works and motion pictures that could not be answered in any other way.<sup>1</sup> TDM can generate new knowledge about the works themselves, but it can also yield broader insights about art, language, culture, and society. And TDM of in-copyright works allows the answers to research questions to be accurate and fully representative of the diversity of our world today, rather than the biased and inaccurate picture that emerges from works published before 1926.

Opponents do not deny the value of TDM. Instead, their primary concern with the proposed exemption is that it is too broad. Petitioners do not object to narrowing their proposal so long as the core purposes of advancing research and teaching are preserved. For example, petitioners would not object to excluding computer programs and limiting eligible works to those lawfully obtained by the researcher or their affiliated institution; limiting beneficiaries of the exemption to researchers affiliated with nonprofit libraries, archives, museums, or institutions of higher education; limiting the purpose of the exemption to scholarly research and teaching; limiting access only to other researchers for project collaboration and verification of research results; and requiring the use of reasonable security measures.

Petitioners’ proposed use is a noninfringing, fair use under the holdings of *Authors Guild, Inc. v. HathiTrust*<sup>2</sup> and *Authors Guild v. Google, Inc.*<sup>3</sup> Opponents’ arguments to the contrary seek to narrow if not relitigate those holdings. The use petitioners propose involves no public access to the corpora, and corpora that any researcher realistically could assemble would be much smaller than the collections involved in either of these cases. The use here is modest by comparison to those found fair in both cases and equally transformative if not more so.

Finally, the alternatives offered by opponents are not viable. Despite the detailed research agendas set out by proponents, opponents do not explain how a single one of the questions proponents seek to study could be satisfactorily addressed using existing collections. HathiTrust and commercial databases are too limited and siloed to answer these questions and are available only to a small subset of researchers. While opponents characterize the drawbacks of optical character recognition (“**OCR**”) and screen capture as “mere inconveniences,” this Office’s own prior conclusions and

---

<sup>1</sup> Throughout, this Reply Comment uses “petitioners” to refer to the signatory organizations seeking the exemption and “proponents” to refer to the researchers and organizations who submitted letters in support of the petition.

<sup>2</sup> 755 F.3d 87 (2d Cir. 2014).

<sup>3</sup> 804 F.3d 202 (2d Cir. 2015) (“**Google Books**”).

proponents' letters demonstrate that these methods are too resource-intensive, poor in quality, and error-ridden to serve as alternatives.

Petitioners have made the required showing, and the Office should recommend granting the exemption.

#### **ITEM D. TECHNOLOGICAL PROTECTION MEASURES AND METHODS OF CIRCUMVENTION**

In our initial comment, we proposed that the technological protection measures and methods (“**TPMs**”) of circumvention covered include those that apply to literary works distributed electronically and to motion pictures.<sup>4</sup> The Office would be well justified in accepting this proposal, as it is either identical to, or a modest extension of, language in recently granted exemptions for these classes of works.<sup>5</sup> Indeed, the only difference from currently granted exemptions is the inclusion of Ultra HD Blu-ray discs protected by AACS2.<sup>6</sup>

The opposition comments offer no meaningful reasons why our proposal regarding TPMs and methods of circumvention should be denied. Two opposition comments fail to address the TPMs in petitioner’s proposal.<sup>7</sup> The Joint Creators and Copyright Owners (“**Joint Creators**”) devote a single sentence to characterizing the TPMs included in the proposed exemption as unnecessarily broad,<sup>8</sup> but do not name any specific access control or method of circumvention that should be omitted and do not address the fact that the Librarian has approved similarly scoped exemptions in the past.<sup>9</sup> The DVD Copy Control Association (“**DVD CCA**”) and Advanced Access Content System Licensing Administration (“**AACS LA**”) likewise make no argument against petitioner’s

---

<sup>4</sup> Authors Alliance et al. Class 7(a) & 7(b) Initial at 8 (“**Initial Comment**”).

<sup>5</sup> See 37 C.F.R. § 201.40(b)(3) (“[l]iterary works, distributed electronically, that are protected by technological measures that either prevent the enabling of read-aloud functionality or interfere with screen readers or other applications or assistive technologies . . .”); *id.* at § 201.40(b)(1) (“[m]otion pictures (including televisions shows and videos) . . . where the motion picture is lawfully made and acquired on a DVD protected by the Content Scramble System, on a Blu-ray disc protected by the Advanced Access Content System, or via a digital transmission protected by a technological measure . . .”).

<sup>6</sup> Acting Register of Copyrights, Section 1201 Rulemaking: Seventh Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention, Recommendation of the Acting Register of Copyrights, at 40 (Oct. 5, 2018) (“**2018 Recommendation**”).

<sup>7</sup> One opposition comment does not address Item D. Software and Information Industry Association Class 7(b) Opp’n. Another expresses support for TPMs generally and disapproval of circumvention, but does not address petitioner’s specific proposed language. Association of American Publishers (“**AAP**”) Class 7(b) Opp’n at 1–2.

<sup>8</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 2.

<sup>9</sup> *Supra* note 5.

proposal except for the blanket assertion that Petitioners do not need access to “high quality content,”<sup>10</sup> an argument fully rebutted by our initial comment.<sup>11</sup>

Nonetheless, petitioners do not object to narrowing the range of TPMs subject to the exemption. While the literary works sub-class should remain unqualified, for motion pictures, we do not object to modifying our original proposal to exclude AACs2. Nor would we object to limiting TPMs protecting digital transmission of works to those protecting works available for digital download. We discuss narrowing the exemption language further in Item E.1 below.

#### **ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGEMENT USES**

As the record supplied in the first round demonstrates, researchers are adversely affected by section 1201 in their efforts to conduct TDM on in-copyright literary works and motion pictures. At present, section 1201 restricts the inquiry into what literature and film tell us about art, language, culture, and society to only what public domain works and a handful of incomplete and limited collections allow researchers to ask. This cordons off most literary works of contemporary relevance and virtually all motion pictures. Moreover, that limitation reencodes and perpetuates biases reflected in public domain works. Broadening that inquiry to include certain in-copyright works is a noninfringing use protected by section 107. Further, the proposed exemption satisfies the conditions for an exemption under section 1201.

Petitioners do not object to a more narrowly framed exemption that allows valuable TDM research to go forward while addressing concerns raised by opposition commenters. In this section, we explain those limitations we are willing to accept and those that are not reasonable. We then address opponents’ attempt to relitigate *HathiTrust* and *Google Books*, and explain how the use contemplated by the proposed exemption is a clear fair use under those holdings. Finally, we explain why the alleged alternatives offered by opponents are not viable, and how section 1201 causes the adverse effect petitioners seek to redress here, notwithstanding what other adverse effects researchers may face due to contractual restrictions.

##### **1. The original exemption request is fully supported by law, but petitioners would not object to a narrower exemption provided that the core purposes of the exemption are preserved.**

Opponents do not deny the value of the research that proponents seek to carry out but contend the exemption as proposed is too broad with respect to the exemption’s classes of works, beneficiaries, and permissible uses.<sup>12</sup> Petitioners would not object to a narrower exemption provided that the scope of the exemption remains broad enough to permit their teaching and research to go forward. This section addresses opponents’ arguments pertaining to the scope of the exemption. We first address those that we believe can be accommodated by narrowing the exemption request in a way

---

<sup>10</sup> DVD CCA & AACs LA Class 7(a) Opp’n at ii.

<sup>11</sup> Initial Comment at 14–15.

<sup>12</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 2.

that does not undermine its purpose. We set out a possible revised exemption and then proceed to describe each of the proposed changes and how they answer opponents' objections. We then explain why opponents' remaining arguments regarding the scope of the exemption are without merit.

**a. A more narrowly tailored exemption can address the bulk of opponents' concerns while still enabling the exemption's core purpose.**

As explained further below, a narrowed exemption that addresses concerns raised by opponents while allowing circumvention to enable the research activities described in the petition and supporting letters would read as follows:

**Proposed Class 7(a) (revised):** Motion pictures, where the motion picture is lawfully made and obtained on a DVD protected by the Content Scramble System, on a Blu-ray disc protected by the Advanced Access Content System, or via a digital transmission protected by a technological protection measure, 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.

**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 narrowed exemption fully addresses opponents' concerns regarding the classes of works subject to the exemption, the beneficiaries who may use the proposed exemption, and the permissible uses and purposes of the exemption. We explain each of the relevant limitations in turn.

**i. Limiting the class of works to exclude computer programs and clarifying that the works must be lawfully obtained.**

Petitioners would not object to two limitations to the class of works subject to the exemption. First, petitioners are willing to exclude computer programs from the literary works subject to the exemption. As is clear from the initial comment and supporting letters, Class 7(b) focuses on works of literature and other writings. Second, petitioners would not object to changing the exemption to

embrace works that are lawfully obtained—rather than lawfully accessed—by the researcher or their affiliated institution. These changes resolve opponents’ concerns about the exemption extending to circumstances that were never contemplated by exemption proponents.

As is clear from the letters supporting the petition, the literary works at the heart of the petition are those distributed as e-books.<sup>13</sup> Nonetheless, the Association of American Publishers (“AAP”) expresses concern that the exemption would extend to “scientific databases or gaming software[.]”<sup>14</sup> Petitioners note that past granted exemptions have relied on context rather than an express exclusion to clarify the literary works subject to the exemption.<sup>15</sup> To avoid doubt, however, petitioners are willing to exclude computer programs. AAP also seeks the exclusion of “databases,” a term not defined in the Copyright Act and for which rightsholders have sought *sui generis* property rights in the past.<sup>16</sup> Given the indeterminacy of a term that could refer to any number of compilations of works, excluding “databases” as a whole is inappropriate. However, petitioners do not object to clarifying that the exemption does not allow circumvention of access controls on compilations that have been compiled specifically for TDM purposes.

Petitioners also are willing to clarify that both the literary works and motion pictures within the scope of the exemption are those that have been lawfully obtained by the researcher or their affiliated institution. We narrow the proposed exemption this way to address two concerns raised by opponents. First, AAP contends that the exemption does not require that the researcher, rather than some unknown third party, have lawful access to the work.<sup>17</sup> Along similar lines, Joint Creators invoke the specter of the periodical database Sci-Hub and the possibility of “other academics . . . devising legal theories as to why academics may use these infringing copies.”<sup>18</sup> The

---

<sup>13</sup> See, e.g., Letter from Andrew Piper, Appendix I to Initial Comment, at 3 (“[E]-books present one of the greatest opportunities for direct access to high-quality text files without requiring a massive time investment.”).

<sup>14</sup> AAP Class 7(b) Opp’n at 3.

<sup>15</sup> See Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, 83 Fed. Reg. 54,010, 54,029 (Oct. 26, 2018) (granting an exemption for “[l]iterary works, distributed electronically” that are distributed with TPMs that prevent or interfere with the use of assistive technologies, without excluding software or databases) (“**2018 Final Rule**”).

<sup>16</sup> See, e.g., Database Investment and Intellectual Property Antipiracy Act of 1996, H.R. 3531, 104th Cong. 2d Sess. (1996).

<sup>17</sup> AAP Class 7(b) Opp’n at 2.

<sup>18</sup> Joint Creators 7(a) & 7(b) Opp’n at 4–5 (citing Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 U.C. Davis L. Rev. 893 (Dec. 2019)). Joint Creators also state that the initial comment discussed the Sci-Hub database. This is not accurate. The reference to Sci-Hub appeared in a supporting letter, and specifically made the point that section 1201’s prohibition against circumvention even for noninfringing uses creates the perverse

fear is ill-founded, but the clarification that researchers or their institutions must lawfully obtain the work puts it to rest.

Second, clarifying that a researcher must lawfully obtain the work should resolve a concern that Joint Creators raise with respect to researchers renting motion pictures and then making permanent copies for TDM research purposes.<sup>19</sup> Again, the letters supporting the exemption make clear that “rent-to-own” is not researchers’ intention for motion pictures or literary works. For example, David Bamman stated that the fast and accurate way for a researcher to create a TDM dataset “would be to *buy* DVDs and use software to rip the film from that medium.”<sup>20</sup> Similarly, 100% of respondents to a survey of members of the Association for Computers and the Humanities (“ACH”) said that circumventing TPMs “on legally *purchased* ebooks would be useful for their research.”<sup>21</sup> Finally, Henry Alexander Wermer-Colan explained that researchers want to buy copies of works for consumptive purposes even when examining those same works as part of TDM dataset.<sup>22</sup> Petitioners do not object to clarifying that the exemption does not allow the retention of permanent copies of rented works.

We note, however, that “lawfully obtained” should not be construed to require outright ownership of digital works subject to the exemption. Copyright owners and digital intermediaries frequently characterize the distribution of digital works as the licensing of the work rather than a sale.<sup>23</sup> Even when a user pays a one-time fee for permanent access to a digital work, whether they own the work

---

incentive to turn to sources like Sci-Hub. Letter from Rachael Samberg & Tim Vollmer, Appendix K to Initial Comment, at 3–4.

<sup>19</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 4.

<sup>20</sup> Letter from David Bamman, Appendix B to Initial Comment, at 3 (emphasis added).

<sup>21</sup> Letter from the ACH, Appendix A to Initial Comment, at 1 (emphasis added). *See also* Letter from the Data-Sitters Club, Appendix D to Initial Comment, at 3 (declaring its interest in purchasing digital editions of books rather than relying on less expensive print copies to be manually disassembled and scanned); Letter from Jes Lopez, Appendix H to Initial Comments, at 2 (“An exemption to § 1201 would enable me to quickly extract the necessary text from purchased eBooks[.]”); Letter from Ted Underwood, Appendix N to Initial Comment, at 3 (“[I]f I could also mine the texts of volumes I had purchased, I could develop supplementary collection . . . to fill gaps in the academic libraries that created HathiTrust.”).

<sup>22</sup> Letter from Henry Alexander Wermer-Colan, Appendix P to Initial Comment, at 3 (“Researchers want to buy books and read them, the better to understand what they see when they look at these texts at scale through the lens of computational algorithms.”).

<sup>23</sup> See, e.g., Amazon, *Kindle Store Terms of Use* (Mar. 15, 2016), <https://www.amazon.com/gp/help/customer/display.html?nodeId=200771440> (last visited Mar. 10, 2021) (“Kindle Content is licensed, not sold, to you by the Content Provider.”).

is open to question.<sup>24</sup> The exemption should not require researchers or their institutions to own copies of digital works when publishers and intermediaries insist that those works are not in fact offered for sale.

**ii. Limiting beneficiaries to researchers affiliated with nonprofit libraries, archives, museums, or institutions of higher education.**

Petitioners do not object to limiting beneficiaries of this exemption to researchers affiliated with nonprofit libraries, archives, museums, or institutions of higher education. An exemption limited to researchers affiliated with qualifying institutions will allow exemption proponents to engage in the research projects and teaching described in their letters while addressing opponents' concerns regarding exploitation of the exemption by commercial actors or bad actors. We first describe the limitation and how it relates to the scope of beneficiaries in previously granted exemptions. We then explain how the limitation addresses the specific objections raised by exemption opponents.

The Office previously has limited exemptions to use by "an eligible library, archives, or museum" as a way of ensuring that an exemption is used for intended, noninfringing purposes.<sup>25</sup> The Office has similarly limited exemptions to "college and university faculty and students[.]"<sup>26</sup> We propose "institutions of higher education" to further narrow the category of colleges and universities. The term draws from the statutory definition set forth in the Higher Education Act, as amended in 20 U.S.C. § 1001, and disclaims for-profit institutions. The statute defines an institution of higher education 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 post-secondary 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.<sup>27</sup>

This limitation addresses the concern raised by some opponents that the exemption is not limited to nonprofit organizations.<sup>28</sup> Even for-profit universities are excluded. Second, the narrower category of institutions who have in-house researchers limits the exemption to institutions accustomed to and capable of using reasonable security measures and data management practices

---

<sup>24</sup> See AARON PERZANOWSKI & JASON SCHULTZ, THE END OF OWNERSHIP: PERSONAL PROPERTY IN THE DIGITAL ECONOMY 75 (2016) (noting courts' struggle to distinguish sales from licenses even for digital works distributed on physical media).

<sup>25</sup> See, e.g., 2018 Recommendation at 255 (software preservation).

<sup>26</sup> Id. at 88 (motion pictures used for educational purposes).

<sup>27</sup> 20 U.S.C. § 1001(a).

<sup>28</sup> Joint Creators Class 7(a) & 7(b) Opp'n at 5; see also DVD CCA & AACS Class 7(a) Opp'n at 11 (claiming there is no context by which one could determine who qualifies as a researcher).

to secure sensitive data. Our initial comment referred to these practices, and we describe them in further detail below.<sup>29</sup>

The narrowed exemption further limits beneficiaries to “researchers affiliated” with qualifying institutions. That term is intended to encompass faculty and staff, as well as students working under their supervision. Academic faculty are likely to carry out the vast majority of the TDM research conducted pursuant to the exemption. However, as the letters accompanying the initial comment demonstrate, post-graduate student researchers also have designed meritorious research projects that they could carry out only with the exemption.<sup>30</sup> The exemption creates room for those projects, and for circumvention by an instructor to develop a corpus to teach TDM techniques to students, while still ensuring that the researcher is affiliated with an institution capable of properly securing the research corpus. The requirement of institutional affiliation allays opponents’ concern that “corporate actors” might make use of the exemption.<sup>31</sup>

### **iii. Limiting the exemption to the use of TDM methods for scholarly research and teaching.**

Petitioners do not object to clarifying that the exemption is limited to the use of TDM methods for scholarly research and teaching. The purpose of the exemption is exceedingly clear from the initial comment and supporting letters: academic researchers want to assemble corpora of works to gain the specific insights into those corpora that only TDM can yield. Faculty also want to teach their students TDM techniques—“longstanding methodological pillars” of the humanities and social sciences—using works that are more relevant, accessible, and useful to those students.<sup>32</sup> Petitioners are willing to make this purpose explicit in the revised exemption.

That limitation resolves concerns that preoccupy exemption opponents but are far removed from the intent of the exemption. The exemption does not seek to allow “profit-seeking individuals and entities to circumvent copyrighted works[.]”<sup>33</sup> The exemption is not intended to pursue “the

---

<sup>29</sup> Initial Comment at 28 (noting that higher education institutions have the technological infrastructure in place to protect TDM collections and are well equipped to help researchers secure their data).

<sup>30</sup> See, e.g., Letter from Jes Lopez, Appendix H to Initial Comment, at 1 (describing a Ph.D. candidate’s intention to use TDM to study neuroatypical narrators in 24 novels should the exemption be granted and explaining how early-career academics lack the resources to carry out these types of research projects without employing TDM techniques).

<sup>31</sup> See Joint Creators Class 7(a) & 7(b) Opp’n at 4 (expressing concern that the exemption could be exploited by “commercial, corporate actors”).

<sup>32</sup> Letter from ACH, Appendix A to Initial Comment, at 1.

<sup>33</sup> AAP Opp’n Class 7(b) at 8. AAP itself seems to acknowledge that the for-profit uses are outside the scope of the exemption, but presses the case nonetheless. See *id.* (noting that the exemption is “seemingly focused on (presumably nonprofit) academic research”).

development of . . . commercially explosive technologies[.]”<sup>34</sup> The exemption does not seek to enable “nonacademic uses[.]”<sup>35</sup> The exemption does not seek to use artificial intelligence to create “new” expressive works that mimic and perhaps even infringe upon works in the corpus from which they [are] derived.”<sup>36</sup> The exemption seeks to enable use of TDM for noncommercial scholarship and teaching. This much is clear from the initial comment and supporting letters, but petitioners do not object to amending the language of the proposed exemption to make it clearer.

#### iv. Requiring researchers to implement reasonable security measures.

Libraries, archives, museums, and academic institutions are well equipped to appropriately manage and secure research corpora. We do not object to this exemption mirroring prior granted exemptions that require reasonable measures to prevent further dissemination of works. Those exemptions included security measures that are reasonable in view of the purpose of the exemption, the persons or entities making use of the work, and the risk of infringement or other harm.<sup>37</sup> Exemptions granted in 2015 and 2018 required measures to “reasonably prevent unauthorized further dissemination of a work.”<sup>38</sup> The context and purpose of the exemption matter. For example, security research conducted pursuant to the exemptions granted in 2015 and 2018 must be “carried out in an environment designed to avoid any harm to individuals or the public” and the information derived may not be “used or maintained in a manner that facilitates copyright infringement.”<sup>39</sup> And organizations availing themselves of the software preservation exemption must implement “reasonable digital security measures” to protect archived software.<sup>40</sup>

Importantly, none of these granted exemptions prescribe the exact security measures that beneficiaries must take. Indeed, the Section 108 Discussion Document from which the software preservation security provision originates warns against such prescription: “The Office believes that attempting to prescribe detailed digital security requirements tailored to each kind of use

---

<sup>34</sup> DVD CCA & AAC Class 7(a) Opp’n at 7. DVD CCA finds support for its contention that the true aim of the exemption is to compete commercially with European-based entities in the letter submitted by Lauren Tilton and Taylor Arnold. *Id.* at 8. But those entities are scholarly initiatives, not commercial ones. The Common Lab Research Infrastructure for the Arts and Humanities (“CLARIAH”) and the Digital Research Infrastructure for the Arts and Humanities (“DARIAH”) are pan-European scholarship networks, not commercial endeavors. *See CLARIAH, About CLARIAH*, <https://www.clariah.nl/en/about> (last visited Mar. 10, 2021); DARIAH-EU, *Dariah in a Nutshell*, <https://www.dariah.eu/about/dariah-in-nutshell/> (last visited Mar. 10, 2021).

<sup>35</sup> AAP Class 7(b) Opp’n at 7.

<sup>36</sup> *Id.* at 9.

<sup>37</sup> *See, e.g.*, 2018 Final Rule at 54,030.

<sup>38</sup> *Id.* at 54,017–18 (criticism and commentary in online courses); *id.* at 54,019 (accessible films in education).

<sup>39</sup> *Id.* at 54,025–26.

<sup>40</sup> *Id.* at 54,023.

would result in an unduly burdensome requirement. Whether an institution's particular digital security measure is 'reasonable' will largely depend upon what measures other institutions of similar size and mission have adopted."<sup>41</sup> A similar approach is appropriate here.

As key research hubs, qualifying institutions regularly manage and secure sensitive research data.<sup>42</sup> Prior rulemakings have found that security measures of libraries and archives are sufficient to house works decrypted pursuant to granted exemptions,<sup>43</sup> and opponents point to no instances of leaks or breaches at institutions that have availed themselves of past exemptions. Research institutions regularly assist researchers in creating data management plans and securing data, as indicated by resources at proponents' institutions.<sup>44</sup> For example, UC Berkeley has a Research Data Management team of staff from libraries and information technology departments to review the management of data with researchers through the lifecycle of their research projects.<sup>45</sup> They help address questions such as who can access data, how securely the data must be stored, and

---

<sup>41</sup> U.S. Copyright Office, Section 108 of Title 17: A Discussion Document of the Register of Copyrights, at 21 (Sept. 2017).

<sup>42</sup> See Stanford U. IT, *Risk Classifications*, <https://uit.stanford.edu/guide/riskclassifications> (last visited Mar. 10, 2021) (describing the various levels of sensitivity for Stanford research data).

<sup>43</sup> 2018 Final Rule at 54,023.

<sup>44</sup> See Berkeley Research Data Mgmt., *Data Mgmt. Plans*, <https://researchdata.berkeley.edu/services/data-management-plans> (last visited Mar. 10, 2021); Stanford Libraries, *Data Mgmt. Services*, <https://library.stanford.edu/research/data-management-services> (last visited Mar. 10, 2021); U. of Wis., Madison, *Research Data Services*, U. of Wis., Madison, <https://researchdata.wisc.edu/> (last visited Mar. 10, 2021); Wash. State U. Libraries, *Research Data Mgmt.*, Wash. State U. Libraries, <https://libguides.libraries.wsu.edu/rdmllibguide/dmp> (last visited Mar. 10, 2021); U. Chi. Library, *Research Data Mgmt.*, U. Chi. Library, <https://guides.lib.uchicago.edu/datamanagement/storing> (last visited Mar. 10, 2021); Mich. State U. Libraries, *Research Data Mgmt. Guidance*, Mich. State U. Libraries, <https://lib.msu.edu/rdmg/resources/> (last visited Mar. 10, 2021); McGill Library, *Research Data Services*, <https://www.mcgill.ca/library/services/data-services/organizing> (last visited Mar. 10, 2021); Loyola U. Chi., *Off. of Research Services*, Loyola U. Chi., <https://www.luc.edu/ors/additionalservices.shtml> (last visited Mar. 10, 2021); Emory U., *Research Data*, <https://researchdata.emory.edu/plan/index.html> (last visited Mar. 10, 2021); U. of Richmond Boatwright Library, *Data Mgmt.*, <https://libguides.richmond.edu/datamanagement> (last visited Mar. 9, 2021); U. Ill. Library, *Data Mgmt. Overview*, <https://guides.library.illinois.edu/introdata/overview> (last visited Mar. 10, 2021); Grambling State U., *Instructional Excellence Resources*, <https://www.gram.edu/research/resources/> (last visited Mar. 10, 2021); Cornell U., *Research Data Mgmt. Service Group*, <https://data.research.cornell.edu/> (last visited Mar. 10, 2021); Temple U. Libraries, *Research Data Mgmt.*, <https://guides.temple.edu/datamgmt> (last visited Mar. 10, 2021).

<sup>45</sup> Berkeley Research Data Mgmt., *Data Mgmt. Plans*, <https://researchdata.berkeley.edu/services/data-management-plans> (last visited Mar. 10, 2021).

what analysis will be performed on the data.<sup>46</sup> Across institutions, these data management resources stress the importance of securing sensitive types of data, creating access controls for researchers and collaborators, and disseminating research results responsibly.<sup>47</sup> Further, institutions have data storage options that can restrict data access to specific accounts of approved researchers affiliated with the institutions and apply granular controls to data depending on the level of sensitivity.<sup>48</sup>

Opponents also hold out the systems in place at HathiTrust and Google as the minimum standard for security practices,<sup>49</sup> but this ignores that reasonable security is a function of institution size, degree of public access, and type of research. Also, it is a mischaracterization to say that institutions' ability to secure data is vastly different than that of HathiTrust. The HathiTrust Research Center is a collaboration of Indiana University and the University of Illinois and is supported by Indiana University's libraries and information technology teams.<sup>50</sup> Where HathiTrust's security measures exceed what an individual researcher or institution may use, it is because the scale, scope, and type of access merit heightened caution. Similarly, Google Books' security is appropriate for its status as a global search engine and snippet-provider for a collection of 40 million titles.<sup>51</sup> It would stifle valuable research to require an individual researcher to marshal the same resources as a multinational corporation to study a much smaller corpus of works to which the public has no access.

Just as past exemptions grounded their security inquiry in the overall context of the secondary use, the security measures used by researchers affiliated with qualifying institutions are reasonable given other limitations proposed here to minimize the risk of infringement. Researchers at qualifying institutions are well-situated to appropriately manage and secure their research corpora in a way that limits access to only collaborators and peer reviewers seeking to validate research

---

<sup>46</sup> *Id.*

<sup>47</sup> See, e.g., U. Chi. Library, *Research Data Mgmt.*, U. Chi. Library, <https://guides.lib.uchicago.edu/datamanagement/storing> (last visited Mar. 10, 2021).

<sup>48</sup> See e.g. Berkeley Research Data Mgmt., *Active Data Mgmt.*, <https://researchdata.berkeley.edu/services/active-data-management> (last visited Mar. 9, 2021); Stanford U. IT, *Risk Classifications*, <https://uit.stanford.edu/guide/riskclassifications> (last visited Mar. 10, 2021); U. Chi. Center for Research Informatics, *CRI Research Data Storage*, U. Chi. Center for Research Informatics, <https://cri.uchicago.edu/storage/> (last visited Mar. 9, 2021); Emory U., *Data Storage Resources at Emory*, <https://researchdata.emory.edu/organize/resources.html> (last visited Mar. 10, 2021).

<sup>49</sup> See AAP Class 7(b) Opp'n at 6.

<sup>50</sup> See Indiana U. Pervasive Tech. Inst., *HathiTrust Research Center*, Indiana U. Pervasive Tech. Inst., <https://pti.iu.edu/centers/htrc/index.html> (last visited Mar. 10, 2021).

<sup>51</sup> See Haimin Lee, *15 Years of Google Books*, The Keyword Google Blog (Oct. 17, 2019), <https://www.blog.google/products/search/15-years-google-books/> (last visited Mar. 10, 2021).

findings. By imposing a reasonable security requirement, the Office can balance the interests of valuable research and protection of copyrighted works.

**v. Restricting access to the corpus to other researchers for purposes of collaboration and replication of research or verification of research results.**

Petitioners are willing to expressly limit access to research corpora to other researchers for the limited purposes of collaboration on specific research projects and verification of research results. Simply put, researchers do not seek to make their research corpora publicly available. As explained in the initial comment, the use covered by the exemption does not include displaying any portion of any works in a research corpus to the public; not even a snippet.<sup>52</sup> To the extent a researcher seeks to include a quote or excerpt of a work to illustrate the insight gained from applying TDM methods to a research corpus, they do not need this exemption to do so.

Joint Creators acknowledge as much but still fault the exemption for not expressly prohibiting researchers from making portions of decrypted works publicly available.<sup>53</sup> Joint Creators also imagine a deficiency in the exemption’s refusal to announce which “deemed lawful” uses that may accompany TDM research findings but do not depend on the exemption (such as excerpting a movie to critique it) are allowed and which are not.<sup>54</sup> If a literary critic uses assistive technology to read an e-book, that exemption for assistive technologies does not become impermissibly broad because it does not declare whether the critic may also quote from the work in the resulting article. Similarly, given that other exemptions already exist for excerpting motion pictures, Joint Creators’ preoccupation is misplaced.<sup>55</sup> Nonetheless, petitioners are willing to narrow the exemption to set opponents’ minds at ease.

While petitioners are willing to disclaim public access to decrypted works or use of corpora for purposes unrelated to their research, academic researchers must be able to provide limited access to their corpora to one another for the exemption to achieve its purpose. Digital humanities researchers frequently collaborate on research projects across institutions.<sup>56</sup> Further, like

---

<sup>52</sup> Initial Comment at 17, 23 n.120 (“the use covered by the exemption does not include displaying snippets of text to the general public”). For this reason, DVD CCA’s argument about prior exemptions allowing only short clips of circumvented audiovisual works to be displayed is beside the point. DVD CCA & AACSB LA Class 7(a) Opp’n at 9.

<sup>53</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 5.

<sup>54</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 5; *see also* DVD CCA & AACSB LA Class 7(a) Opp’n at 8–9 (similar).

<sup>55</sup> See, e.g., 2018 Rule at 54,018 (allowing circumvention to add short portions of motion pictures to non-fiction multimedia e-books).

<sup>56</sup> See, e.g., Letter from Hoyt Long, Appendix G to Initial Comment, at 3 (expressing interest to engage in a collaborative research project between the University of Chicago, UC Berkeley, and McGill University).

colleagues in medicine and other fields, academic researchers in the digital humanities have grown increasingly aware of the importance of replicability of research results.<sup>57</sup> Replication is “an independent repetition of an earlier, published study, using sufficiently similar methods (along the appropriate dimensions) and conducted under sufficiently similar circumstances.”<sup>58</sup> That repetition requires the peer reviewer to access the data supporting research findings. The exemption thus allows limited access to researchers affiliated with other qualifying institutions for that limited purpose and for the purpose of collaboration. That access poses no threat to content owners’ interests given that the exemption requires the use of security measures capable of limiting access to research corpora to those stated purposes.

**b. Other objections to the scope of petitioners’ exemption request are without merit.**

While petitioners are willing to narrow the exemption to address the concerns raised above, opponents’ two remaining objections to the scope of the exemption are without merit. Those objections relate to the range of computational activities that fall within the category of TDM techniques, and the combination of computational research findings with conclusions drawn from simply reading or viewing works in the corpora. We address these objections in turn.

First, opponents contend that the range of techniques covered by TDM is impermissibly broad. For example, Joint Creators argue that TDM can refer to “anything done by any person who uses a computer.”<sup>59</sup> However, TDM as a concept has proven sufficiently clear and administrable to be the subject of both copyright limitations and anticircumvention exemptions in other jurisdictions. Article 3 of the European Union’s Digital Single Market Copyright Directive provides for a TDM exemption, and the recitals do not limit “text and data mining” beyond clarifying that the term refers to “automated computational analysis of information in digital form, such as text, sounds, images or data[.]”<sup>60</sup> Japan’s copyright law also has proven capable of implementing an exemption for “data analysis” without a closed list of eligible methods.<sup>61</sup> TDM involves diverse methods.<sup>62</sup>

---

<sup>57</sup> See Rik Peels, *Replicability and Replication in the Humanities*, Res. Integrity and Peer Rev. 4, 2 (2019), <https://doi.org/10.1186/s41073-018-0060-4>.

<sup>58</sup> *Id.*

<sup>59</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 2 n.1.

<sup>60</sup> Directive 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market, 2019 O.J. (L 130), Recital 8 & Art. 3.

<sup>61</sup> See Copyright Law of Japan, Ch. 2, Art. 30-4, <https://www.cric.or.jp/english/clj/cl2.html#art30> (defining “data analysis” as “the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data”).

<sup>62</sup> See Eleanor Dickson et al., *HathiTrust Research Center User Requirements Study White Paper*, at 7–9 (Mar. 9, 2018), <http://hdl.handle.net/2142/99157>.

New ones will continue to emerge. Petitioners are thus unwilling to artificially constrain eligible methods here.

Second, opponents object to researchers doing anything other than computational analysis as part of research projects that involve analytical work enabled by the exemption. AAP complains that while the Data-Sitters Club may include “computational elements” in its analysis of how a book series treats religion, race, adoption, divorce, and disability, the overall project “does not sound merely computational in nature.”<sup>63</sup> The overall project isn’t “merely computational,” nor should it be. An article containing nothing but the raw output of statistical analyses or interpretive models would be unconvincing, unenlightening, and borderline unreadable. The connection of insights drawn from computational analysis with insights drawn from reading individual works illustrates, grounds, and validates the research findings that TDM makes possible. But circumvention and computational analysis are not necessary to engage directly with literature or film, even if the insights drawn from that engagement are paired with and enhanced by insights drawn from computational analysis. Declaring conventional, long-established methods of the humanities off-limits to a digital humanities scholar using the exemption would needlessly constrict the value and impact of their research.

## **2. TDM for the purpose of scholarly research and teaching is a fair use.**

The Second Circuit’s decisions in *HathiTrust* and *Google Books* are faithful applications of fair use precedent, and the use at the center of this proposed exemption fits comfortably within their holdings. Opponents’ argument that the proposed exemption asks the Office to “break new ground” on fair use is an attempt to cabin the holdings of those cases, if not relitigate their outcomes. This proceeding is not the appropriate venue for that advocacy. A straightforward application of *HathiTrust* and *Google Books*, as well as cases that precede and follow them, yield the conclusion that TDM for academic research and teaching is a transformative fair use that does not result in market harm. Indeed, the gap between opponents’ characterization of potential licensing opportunities and the opportunities that actually exist demonstrates both the frailty of their market harm argument and the necessity of the proposed exemption.

- a. ***HathiTrust* and *Google Books* are settled law and support the conclusion that the secondary use enabled by the exemption is fair use.**

Opponents’ fair use arguments turn on two contentions. First, that *HathiTrust* and *Google Books* are the water’s edge of fair use with respect to digitization of multiple works, if not wrongly decided. Second, that the uses enabled by the proposed exemption, even with the limitations described in this reply, fall outside of the cases’ holdings. Neither contention is correct.

First, *HathiTrust* and *Google Books* are faithful applications of a chain of precedent from cases involving digitization of multiple works for transformative purposes. In *HathiTrust*, the Second

---

<sup>63</sup> AAP Class 7(b) Opp’n at 7–8.

Circuit explained how *Kelly v. Arriba Soft Corp.*,<sup>64</sup> *Perfect 10, Inc. v. Amazon.com, Inc.*,<sup>65</sup> and *A.V. ex rel. Vanderhye v. iParadigms, LLC*,<sup>66</sup> all reinforced the conclusion that enabling full-text search of entire works “adds a great deal more to the copyrighted works at issue than did the transformative uses [the court] approved in several other cases.”<sup>67</sup> Like *HathiTrust* and *Google Books*, all of those cases involved creating digital copies of multiple entire works. The court in *Google Books* looked to the same cases again, and concluded that there was “no doubt that the purpose of this copying [to enable full-text searching] is the sort of transformative purpose described in *Campbell*[.]”<sup>68</sup> While *Google Books* may have “test[ed] the boundaries of fair use[,]” it passed the test.<sup>69</sup> Moreover, the factors that required the court to grapple with the boundaries of fair use—the scale and commercial nature of the project, and making snippets publicly viewable—are all beyond what petitioners seek here.

Underlying opponents’ efforts to cast *Google Books* and *HathiTrust* as outliers if not aberrations among fair use cases is the fact that opponents disagree with those cases’ holdings. Either directly or through their affiliated trade associations, opponents participated as plaintiffs or amici supporting the plaintiffs in both cases.<sup>70</sup> But just as the triennial review is not the venue to break new ground on fair use, neither is it the venue to reinterpret and narrow case law. AAP may not “endorse unauthorized systematic scanning of full-text copyrighted works, which it believes to be incompatible with fair use.”<sup>71</sup> But no matter how strongly held, that belief must give way to the case law that actually settled the question. It may be, as AAP notes, that the court in *Fox News Network, LLC v. TVEyes, Inc.* “made a point of stating that it ‘expressed no views on the [search function of the database], neither upholding nor rejecting it.’”<sup>72</sup> However, the court did not decline to do so because the law was unsettled as to whether the search function was fair use. Rather, the court did not address the question because the law was so sufficiently settled, even as applied to audiovisual works rather than literary works, that Fox did not bother to challenge it.<sup>73</sup>

---

<sup>64</sup> 336 F.3d 811, 819 (9th Cir. 2003).

<sup>65</sup> 508 F.3d 1146, 1165 (9th Cir. 2007).

<sup>66</sup> 562 F.3d 630, 639–40 (4th Cir. 2009).

<sup>67</sup> *HathiTrust*, 755 F.3d at 97–98.

<sup>68</sup> *Google Books*, 804 F.3d at 217.

<sup>69</sup> *Id.* at 206.

<sup>70</sup> AAP was a plaintiff in *Google Books* and amicus in support of the plaintiffs in *HathiTrust*. The Motion Picture Association of America was amicus in support of the plaintiffs in both *HathiTrust* and *Google Books*.

<sup>71</sup> AAP Class 7(b) Opp’n at 12.

<sup>72</sup> AAP Class 7(b) Opp’n at 4 (quoting *Fox News Network, LLC v. TVEyes, Inc.*, 883 F.3d 169, 182 n.7 (2d Cir. 2018) (alteration AAP’s)).

<sup>73</sup> *TVEyes*, 883 F.3d at 182 n.7.

Petitioners request that the Office take care in evaluating claims that an exemption “breaks new ground” on fair use. In 2003, the Register of Copyrights rightly and carefully concluded that intermediate copying of e-books for use with assistive technologies was within the spirit of section 1201 “and most likely would . . . appear to constitute a fair use of the work.”<sup>74</sup> It took more than a decade for a court to address the question, ultimately agreeing with the Register.<sup>75</sup> Forcing petitioners to wait for that decision in no way would have furthered the purposes of section 1201 or copyright. Regardless, here there is relevant caselaw that establishes that the proposed exemption enables noninfringing uses. AAP’s claim that “[o]nly a single circuit court” has meaningfully addressed the fair use question ignores *HathiTrust* and *Google Books*’ relationship to predecessors like *Kelly*, *Perfect 10*, and *iParadigms*.<sup>76</sup> It also moves the goal posts on what it means to break new ground on fair use by suggesting that multiple circuits must address a question before the ground is broken. That standard would all but end the ability to rely on fair use in section 1201 proceedings. Fair use may be a case-by-case inquiry, but petitioners are entitled to rely on the cases that have been decided.<sup>77</sup>

The proposed exemption falls well within the range of fair uses those cases recognize. Indeed, *Google Books* expressly mentioned TDM as one of the “new forms of research” that *Google Books* and the Google Library Project make possible, citing the research of one of the current exemption

---

<sup>74</sup> Recommendation of the Register of Copyrights in RM 2002-4, Rulemaking on Exemptions from Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, at 70 (Oct. 27, 2003).

<sup>75</sup> *HathiTrust*, 755 F.3d at 103 (holding that “the doctrine of fair use allows the Libraries to provide full digital access to copyrighted works to their print-disabled patrons”); *see also* Register of Copyrights, Section 1201 Rulemaking: Sixth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention, Recommendation of the Register of Copyrights, at 134 (Oct. 8, 2015) (noting the *HathiTrust* decision in concluding that the proponents had presented a compelling case for renewing the exemption for assistive technologies).

<sup>76</sup> AAP Class 7(b) Opp’n at 4.

<sup>77</sup> Despite AAP’s suggestion to the contrary, *id.* at 5, scholar Matthew Sag states unequivocally that reproduction for TDM are fair use under *HathiTrust* and *Google Books*. Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. Copyright Soc’y of the U.S.A. 291, 293–94 (2019) (“Those cases held unambiguously that reproducing copyrighted works as one step in the process of knowledge discovery through text data mining was transformative, and thus ultimately a fair use of those works.”). DVD CCA separately contends that *HathiTrust* and *Google Books* do not apply to motion pictures because books are analog works distributed without TPMs whereas movies are digital works distributed with TPMs. DVD CCA & AACS LA Class 7(a) Opp’n at 12–13. However, those distinctions were not relevant to the courts’ holdings in either case.

proponents.<sup>78</sup> *Google Books* explained how these forms of research enable deep study of “how nomenclature, linguistic usage, and literary style have changed over time[,]” precisely the types of inquiries at issue here.<sup>79</sup> And, again, proponents do not seek to use the exemption for purposes that the court deemed to be on the other side of the fair use boundary in *TVEyes*. The court found that the ability to watch and download unlimited ten-minute clips of television news coverage via *TVEyes*’ “Watch function,” along with the commercial nature of the project, undermined the defendant’s fair use claim despite the secondary use’s modestly transformative character.<sup>80</sup> Essentially, the court found that *TVEyes* was a commercial service that republished some of the copyright owner’s content unaltered.<sup>81</sup> Here, by contrast, the core use enabled by the exemption is noncommercial and does not involve republishing any portion of the works included in a research corpus.

These limitations, among others, bring the proposed use squarely within fair use under the holdings of *HathiTrust* and *Google Books*. The use is noncommercial and transformative. No snippet or other portion is made available to the public directly from the corpus. And the institutions use reasonable security measures to guard against infringement.

**b. TDM for scholarly research and teaching is a noninfringing fair use under the four statutory factors.**

TDM for scholarly research and teaching is a well-established transformative use. The other statutory factors support the conclusion that the secondary use proponents seek to engage in is a fair use. Moreover, opponents’ assertion of market harm falls short and, if anything, only further demonstrates the need for the exemption.

**i. TDM for academic research and teaching is a transformative use expressly favored by section 107.**

Both the text of section 107 and the guidance that emerges from its application to specific cases support the conclusion that TDM for scholarly research and teaching is a transformative, fair use. First, in assessing the “character” of the secondary use, courts consider the specific examples set forth in section 107’s preamble.<sup>82</sup> Those uses include scholarship, research, and teaching.<sup>83</sup> Those are precisely the uses enabled by the exemption when narrowed to proponents’ core purposes.

---

<sup>78</sup> *Google Books*, 804 F.3d at 209 & n.6 (citing David Bamman & David Smith, *Extracting Two Thousand Years of Latin from a Million Book Library*, *J. Computing & Cultural Heritage* 5 (2012), 1–13).

<sup>79</sup> *Google Books*, 804 F.3d at 209.

<sup>80</sup> *TVEyes*, 883 F.3d at 181.

<sup>81</sup> *Id.* at 178.

<sup>82</sup> *iParadigms*, 562 F.3d at 638.

<sup>83</sup> 17 U.S.C. § 107.

Further, those uses are highly transformative. From *Campbell* onward, courts have recognized that uses that comment on or provide information about the original work most clearly satisfy the inquiry into transformative use.<sup>84</sup> And as with *iParadigms*, *HathiTrust*, and *Google Books*, the purpose of the circumvention and the copying that the exemption enables is to “make available significant information *about those books*.<sup>85</sup> The exemption’s aim, as clear from the supporting letters and the limitations discussed above, is a paradigmatic transformative purpose.

Aside from their concerns with the breadth of the exemption, opponents’ arguments against a finding of transformative purpose rest largely on three premises. First, that there are no “limiting principles” on the types of computational research the exemption allows. Second, that researchers and perhaps others would have access to the entirety of the works they circumvent. And finally, that the results of computational analysis may be combined with the results of “traditional textual analysis” that does not rely on computational research and may include reading, viewing, and even quoting the underlying works.<sup>86</sup> None of these premises disturb the conclusion that factor one supports a finding of fair use.

First, placing artificial limitations on eligible methods of computational analysis would be unwarranted even if opponents provided any information on what guardrails they deem appropriate. From the citation to Benjamin Sobel’s article, the concern seems to be that “machines, too, may read for expressive content.”<sup>87</sup> In a sense, this is true. TDM is not a sterile inquiry into whether contemporary novels have more or fewer adverbs than they used to. In one recent example, Andrew Piper developed and applied a model trained on the *Confessions of St. Augustine* and other works to examine the “conversional force” of novels.<sup>88</sup> That inquiry arguably required machines to “read” the expressive content of the works. However, the nub of the

---

<sup>84</sup> *Campbell*, 510 U.S. at 580 (“For the purposes of copyright law, the nub of the definitions . . . is the use of some elements of a prior author’s composition to create a new one that, at least in part, comments on that author’s works”); see also *Google Books*, 804 F.3d at 215–16 (“copying from an original for the purpose of criticism or commentary on the original or provision of information about it, tends most clearly to satisfy *Campbell*’s notion of the ‘transformative’ purpose”).

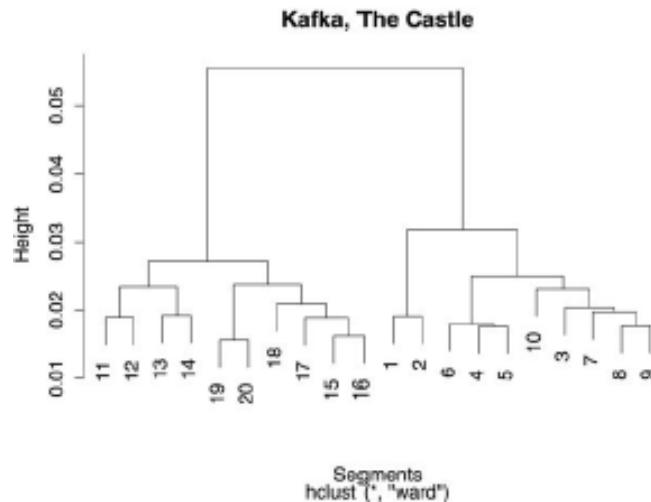
<sup>85</sup> *Google Books*, 804 F.3d at 217 (emphasis in original). For this reason, Joint Creators’ unexplained assertion that *Campbell* runs contrary to a finding that the research purposes contemplated here are transformative is incorrect. Joint Creators Class 7(a) & 7(b) Opp’n at 3 n.6. DVD CCA and AACSB LA’s attempt to avoid this clear chain of precedent by recasting the proposed exemption as one for “space shifting” is similarly wide of the mark. DVD CCA & AACSB LA Class 7(a) Opp’n at 10.

<sup>86</sup> AAP Class 7(b) Opp’n at 7.

<sup>87</sup> *Id.* at 8 (citing Benjamin L. W. Sobel, *Artificial Intelligence’s Fair Use Crisis*, 41 Colum. J. L. & Arts 45, 57 (2017)).

<sup>88</sup> Andrew Piper, *Novel Devotions: Conversional Reading, Computational Modeling, and the Modern Novel*, 46 New Literary History 63 (2015).

transformativeness inquiry is whether the result of that reading “is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn.”<sup>89</sup> That is unquestionably the case here. Here, for example, is Piper’s representation of *The Castle* by Franz Kafka:<sup>90</sup>



In no way is this a substitute for the expressive content of the original work. The fair use inquiry does not and should not require a crabbed definition of eligible TDM methods that would prohibit a digital humanities scholar from studying the expressive content of works in the corpora.<sup>91</sup>

To opponents’ second point, proponents need access to the entirety of the works they place in their corpora but do not seek to provide access to others but for the purposes of collaboration or the replication and verification of research findings. Access to the entirety of the work is necessary for the exemption to achieve its purpose. Computational analysis of anything less than entire works would yield inaccurate and unreliable findings. However, the exemption does not seek access “without any limitations on use of the expressive content.”<sup>92</sup> Access is only for scholarly research and teaching. And, again, the exemption does not contemplate providing access to the corpora to anyone but other researchers, and then only for purposes of collaboration or the replication and verification of research findings.

---

<sup>89</sup> *HathiTrust*, 755 F.3d at 97.

<sup>90</sup> Piper, *supra* note 88, at 89.

<sup>91</sup> To the extent that AAP’s objection to the analysis of expressive content is that the digital humanities scholar is not paying to “devour” texts while the “conventional literary hermeneut” must pay, this argument is a canard. AAP Class 7(b) Opp’n at 8 (quoting Sobel at 82). As discussed in the initial comment and above, exemption proponents are willing to pay for the works they use. However, even if the works they sought were available in usable formats, they should not have to pay twice.

<sup>92</sup> AAP Class 7(b) Opp’n at 7.

Finally, opponents raise the alarm that digital humanities scholars seek to pair computational analysis with more traditional analysis.<sup>93</sup> They do, and nothing about that pairing disturbs the conclusion that reproduction for purposes of TDM is transformative. Of course digital humanities scholars will also want to read or watch works they study via computational methods. And they might even want to quote those works in an article discussing the findings of their TDM. But they do not need this exemption to read or watch works they already have lawfully obtained, and they do not need this exemption to quote from those works. Nothing about those uses that in no way rely on this exemption disturbs the conclusion that the contemplated uses that do rely on this exemption are transformative.

**ii. While not dispositive, the nature of the works weighs in favor of fair use.**

The second fair use factor is not dispositive and is seldom considered in isolation.<sup>94</sup> Its consideration is linked to the first factor. In *Google Books*, the second factor favored fair use not because the works were factual, “but because the secondary use transformatively provides valuable information about the original, rather than replicating protected expression in a manner that provides a meaningful substitute for the original.”<sup>95</sup> That is also the case here.

**iii. Copying entire works is consistent with a finding of fair use when that copying is necessary to achieve the transformative purpose.**

Opponents acknowledge that copying entire works is necessary for the exemption to achieve its purpose.<sup>96</sup> Because copying entire works for the purpose of TDM research is not excessive or unreasonable in view of the purpose of the use proposed here, the third factor favors fair use.<sup>97</sup>

Opponents’ argument that factor three does not favor fair use centers on the claim that the exemption allows “sharing” or “redistribution” of the work.<sup>98</sup> It does not. To be clear, the exemption does not contemplate redistribution of the research corpus and other researchers’ access to it is limited to purposes of collaboration or the replication and verification of research findings.

---

<sup>93</sup> *Id.* at 7–8.

<sup>94</sup> *Google Books*, 804 F.3d at 220.

<sup>95</sup> *Id.*

<sup>96</sup> AAP Class 7(b) Opp’n at 9.

<sup>97</sup> See *HathiTrust*, 755 F.3d at 99 (“We have no reason to think that these [entire] copies are excessive or unreasonable in relation to the purposes identified by the Libraries and permitted by the law of copyright.”). The propriety of copying entire works is not limited to the context of digitization or search. See, e.g., *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605 (2d Cir. 2006) (copying entire concert posters reasonable in view of the transformative purpose).

<sup>98</sup> AAP Class 7(b) Opp’n at 9.

#### **iv. The contemplated use does not result in cognizable market harm.**

The bulk of opponents' arguments regarding market harm flow from concerns that the limitations described above adequately address. Those concerns are: use for purposes unrelated to scholarly research and teaching, making the work available to others, the lack of adequate security measures to prevent copyright infringement, and the lack of any compensation for the original work. The remaining arguments amount to a demand for a license for uses that lie outside the scope of the copyright owners' exclusive rights. That demand is not a basis for a cognizable market harm. We address opponents' arguments in turn.

First, opponents raise again a concern about lack of specificity regarding TDM, this time to suggest that the proposed exemption permits "substitutional uses of expressive content," training models to generate output that mimics or even infringes upon works in the corpus, or unspecified commercial enterprises.<sup>99</sup> As is clear from the initial petition and the supporting letters, none of those uses are remotely related to the purposes for which proponents seek the exemption. Proponents seek an exemption only for scholarly research and teaching. Petitioners do not object to limitations on the exemption consistent with that purpose.

Second, opponents raise concerns that research corpora may be aggregated or made publicly available.<sup>100</sup> While researchers must be able to provide access to collaborators or peer reviewers for replication or validation of research results, that is the extent to which researchers seek to provide access to their own corpora. And petitioners do not object to modifying the exemption to clarify this limitation.

Neither do petitioners object to a requirement to use reasonable security measures to safeguard research corpora against infringement. Opponents' contention that institutions cannot maintain adequate security measures echoes their claim in *Google Books* that Google's distribution of digital copies to participating academic libraries was not fair use because those copies exposed plaintiffs to a risk of loss.<sup>101</sup> The court noted that the agreement between Google and the libraries required them to take precautions to prevent dissemination of their digital copies to the public at large, and that any claim that the copies would be vulnerable to hacking was "nothing more than a speculative possibility."<sup>102</sup> The same is true here. Petitioners agree that the use of reasonable security measures is appropriate. Qualifying institutions are capable of putting reasonable security measures in place, as evidenced by the lack of any instances of hacking into the software archives made possible by the 2015 and 2018 rulemakings.

The final concern to be answered by a narrowed exemption is that researchers will be able to engage in "consumptive uses" of works without lawfully obtaining the copy. Again, the exemption

---

<sup>99</sup> *Id.*

<sup>100</sup> Joint Creators Class 7(a) & 7(b) Opp'n at 4.

<sup>101</sup> *Google Books*, 804 F.3d at 228–29.

<sup>102</sup> *Id.* at 229.

does not seek to put the digital humanities scholar in any position more favorable than the conventional hermeneut or ordinary reader when it comes to lawfully obtaining the work. And petitioners are willing to make this condition explicit to address Joint Creators' concern regarding translating temporary rentals into permanent copies.<sup>103</sup> As multiple supporting letters make clear, researchers are willing to buy works in workable formats.<sup>104</sup>

However, the law does not oblige proponents to acquire, in addition to the work, a license for a transformative fair use, as opponents demand. *HathiTrust* and *Google Books* are clear on this. "Lost licensing revenue counts under Factor Four only when the use serves as a substitute for the original and the full-text-search use does not."<sup>105</sup> Even when the rightsholder seeks to license that use, "the availability of licenses for providing unprotected information about a copyrighted work, or supplying unprotected services related to it" does not give the copyright holder "the right to exclude others from providing such information or services."<sup>106</sup> It is of no moment that the Copyright Clearance Center ("CCC") has built a licensed product that enables TDM of scientific, technical, and medical periodicals.<sup>107</sup> "Current doctrine correctly recognizes that copyright owners' willingness to license, control, or monetize a use does not mean that the use is unfair if unauthorized."<sup>108</sup>

Moreover, opponents have not demonstrated that they are willing to license works that exemption proponents seek to use. CCC's TDM product contains scientific journals, not popular fiction. Joint Creators fault proponents for not requesting a license for TDM of motion pictures, but do not even contend that such a license would be available if requested.<sup>109</sup> Instead, Joint Creators point to a license it makes available for "other educational uses" and suggest it "could potentially license the uses at issue."<sup>110</sup> But substantial research has turned up no licensed collection of motion pictures for TDM. It is unclear whether licensing for academic uses would be a priority—or even feasible—for Joint Creators.

---

<sup>103</sup> Joint Creators Class 7(a) & 7(b) Opp'n at 4.

<sup>104</sup> See Letter from David Bamman, Appendix B to Initial Comment, at 3; see also *supra* note 20 (summarizing statements by researchers regarding their intent to purchase works included in research corpora).

<sup>105</sup> *HathiTrust*, 755 F.3d at 100.

<sup>106</sup> *Google Books*, 804 F.3d at 226.

<sup>107</sup> See AAP Class 7(b) Opp'n at 10 (discussing CCC's *RightFind XML for Mining Solution*, <https://www.copyright.com/publishers/rightfind-xml-for-mining-solution/>).

<sup>108</sup> Rebecca Tushnet, *All of This Has Happened Before and All of This Will Happen Again: Innovation in Copyright Licensing*, 29 Berkeley Tech. L.J. 1447, 1486 (2014).

<sup>109</sup> Joint Creators Class 7(a) & 7(b) Opp'n at 6.

<sup>110</sup> Joint Creators Class 7(a) & 7(b) Opp'n at 6; see also DVD CCA & AACS LA Class 7(a) Opp'n at 14–15 (contending that licensing alternatives exist without citing a single example).

And this is the problem. In effect, AAP and Joint Creators are saying a potential licensing opportunity forecloses proponents' chosen research agendas, and that until opponents deem it in their interest to make that potential a reality, proponents must content themselves with what research questions the public domain or, if they are fortunate, HathiTrust, allow them to answer. That is not a credible showing of market harm. Indeed, fair use exists in large part to "facilitate a class of uses that would not be possible if users always had to negotiate with copyright proprietors."<sup>111</sup> That is the case here.

### **3. Section 1201 adversely affects noninfringing TDM research and there is no practical alternative to circumvention that eliminates that harm.**

The alternatives to circumvention proposed by opponents are inadequate. HathiTrust, while a major achievement in archiving, accessibility, and use of literary works, does not meet many researchers' needs. Aside from usability and resource constraints, HathiTrust's collection is limited and available only to researchers affiliated with member institutions. Commercial databases, in addition to being siloed and noninteroperable, have limited collections and do not allow researchers to use computational techniques refined to answer specific research questions. OCR and screen capture would be insufficient alternatives even if researchers had the infinite time and resources to use them in the way opponents suggest. We address the insufficiency of these putative alternatives in turn.

#### **a. HathiTrust's collection is limited and unavailable to many if not most beneficiaries of the proposed exemption.**

Opponents do not contest the point established in the initial comment that the computing resources made available to researchers via HathiTrust data capsules are not sufficient to answer certain research questions.<sup>112</sup> Nor do they contest that using the data capsules makes it challenging to revise code while working with a corpus, as many researchers seek to do.<sup>113</sup> Instead, opponents contend that these limitations are features, not bugs, overlooking that there are reasonable alternative means to secure a research corpus.<sup>114</sup>

Even if opponents were correct, however, HathiTrust's core limitation is that its collection is limited and available only to researchers affiliated with partner libraries. Opponents concede as much, noting that HathiTrust has about 50% of the print holdings of research libraries in North America.<sup>115</sup> In other words, whether HathiTrust has a book available for TDM research is, at best,

---

<sup>111</sup> *Kienitz v. Sconnie Nation LLC*, 766 F.3d 756, 759 (7th Cir. 2014).

<sup>112</sup> Initial Comment at 11 (citing Letter from David Bamman, Appendix B to Initial Comment, at 2).

<sup>113</sup> *Id.* at 11 (citing Letter from Ted Underwood, Appendix N to Initial Comment, at 2).

<sup>114</sup> AAP Class 7(b) Opp'n at 11.

<sup>115</sup> *Id.*

a coin flip. That probability does not work for statistical and computational research findings whose validity turns on the comprehensiveness of the dataset.<sup>116</sup> Moreover, those holdings skew toward academic monographs rather than the works of popular contemporary literature that many exemption proponents seek to study.<sup>117</sup> And HathiTrust includes no motion pictures.

Further, most researchers do not have access to the HathiTrust collection for TDM research purposes. Professor James Clawson at Grambling State University explained that he and colleagues at other historically black colleges and universities do not currently have access to the HathiTrust collection and their institutions are unlikely to become members due to cost constraints.<sup>118</sup> While impressive, HathiTrust's collection in no way solves the problem that has led proponents to seek an exemption from circumvention liability under section 1201.

**b. Commercial databases are too siloed and limited to be meaningful alternatives to circumvention.**

Commercial databases are far more limited than HathiTrust, as is clear from the two examples that AAP cites.<sup>119</sup> CCC's RightFind XML for Mining includes only articles from scientific journals.<sup>120</sup> Gale does include some American fiction in its *Primary Sources* collection, which may also be used with its Digital Scholars Lab.<sup>121</sup> However, *Primary Sources*' collection of American fiction ends with works published in 1920.<sup>122</sup>

Further, commercial databases like Gale's require the use of limited and in some cases proprietary models and algorithms rather than the ones a researcher develops to accomplish a specific research

---

<sup>116</sup> See, e.g., Letter from Ted Underwood, Appendix N to Initial Comment, at 1; Letter from Henry Alexander Wermer-Colan, Appendix P to Initial Comment, at 2–3.

<sup>117</sup> Letter from David Bamman, Appendix B to Initial Comment, at 2; Letter from Henry Alexander Wermer-Colan, Appendix P to Initial Comment, at 2; Letter from Jes Lopez, Appendix H to Initial Comment, at 2; Letter from Ted Underwood, Appendix N to Initial Comment, at 1.

<sup>118</sup> Letter from James Clawson, Appendix C to Initial Comment, at 2.

<sup>119</sup> AAP Class 7(b) Opp'n at 9–10.

<sup>120</sup> CCC, *RightFind XML for Mining Solution* <https://www.copyright.com/publishers/rightfind-xml-for-mining-solution/> (last visited Mar. 10, 2021) (marketing RightFind XML for Mining for biomedical researchers and scientific, technical, and medical publishers); CCC, *Product Sheet: RightFind XML for Mining* <http://www.copyright.com/wp-content/uploads/2016/04/Product-Sheet-RF-XML-for-Mining.pdf> (last visited Mar. 10, 2021) (advertising RightFind XML for Mining as granting access to “full-text scientific, technical, and medical content”).

<sup>121</sup> See Gale, 2019 Gale Primary Source Catalog, <https://www.gale.com/primary-sources/catalog?sort=ascending&pre=true> (last visited Mar. 10, 2021); Gale, Gale Digital Scholar Lab, <https://www.gale.com/primary-sources/digital-scholar-lab> (last visited Mar. 10, 2021).

<sup>122</sup> See Gale, 2019 Gale Primary Source Catalog, *supra* note 121, at 51.

project.<sup>123</sup> As Wermer-Colan explains, these databases offer only “proprietary corpora and text mining tools” that “restrict researchers to relatively random sets of that data in each database.”<sup>124</sup> Thus, there is no way to combine different proprietary databases or use customized research tools across their siloed collections. And, again, motion pictures are not included. These supposed alternatives do not solve TDM researchers’ dilemma so much as clarify it.

**c. OCR and screen capture are not viable alternatives.**

Opponents do not contest the limitations of OCR and screen capture that the record demonstrates thoroughly. Instead, they argue that these limitations are “mere inconveniences.” That argument assumes access to resources beyond those that are available to researchers and qualifying institutions, and ignores inherent flaws in OCR and screen capture that would persist even assuming those resources existed.

The record amply demonstrates that OCR results in a digital file sufficiently error-ridden to require both automated and manual review of the file after scanning is completed.<sup>125</sup> Because TDM requires a high level of precision, this file correction is extraordinarily laborious and time-consuming.<sup>126</sup> Opponents’ dismissal of these hurdles as “mere inconvenience” overlooks that researchers have finite time in which to complete their projects. This is why most work in the digital humanities concerns public domain works.

Further, OCR simply is not a solution for certain types of works. The geometric complexity and variation of ideographic languages makes OCR particularly difficult for works in those languages.<sup>127</sup> Handwriting-style fonts pose similar problems.<sup>128</sup> OCR also cannot reveal

---

<sup>123</sup> See *id.* at 7 (discussing the six text mining methodologies included with Gale’s Digital Scholar Lab).

<sup>124</sup> Letter from Henry Alexander Wermer-Colan, Appendix P to Initial Comment, at 2.

<sup>125</sup> See, e.g., Letter from Henry Alexander Wermer-Colan, Appendix P to Initial Comment, at 1 (explaining how one scanning project involving multiple library departments, a half-dozen staff members, and student works still took three years to digitize 500 books); Letter from ACH, Appendix A to Initial Comment, at 1–2 (a three-year project using OCR was possible only because the scanning commenced one year before the start date).

<sup>126</sup> Letter from Hoyt Long, Appendix G to Initial Comment, at 2; see also Dickson, et al., *HathiTrust Research Center User Requirements Study White Paper*, at 5 (2018), <http://hdl.handle.net/2142/99157> (last visited, Mar. 10, 2021) (relating the comments of a person who works at a digital humanities center that scanning and OCR can be “very time-consuming” and “labor-intensive,” and that the subsequent editing process “can be endless”).

<sup>127</sup> Letter from Hoyt Long, Appendix G to Initial Comment, at 2.

<sup>128</sup> Letter from Data-Sitters Club, Appendix D to Initial Comment, at 2.

provenance or other information embedded in digitized files.<sup>129</sup> OCR simply is not up to the task for most TDM research projects.

Similarly, the Office has recognized that screen capture technology is not a viable alternative for certain purposes related to criticism and commentary.<sup>130</sup> So, too, here. Screen capture results in lower-resolution data than the original medium and strips away important structural markers and metadata.<sup>131</sup> Opponents concede as much.<sup>132</sup> The limitations of screen capture have led to the outright abandonment of research projects of considerable merit that would have taken 10 years of full-time work to complete.<sup>133</sup>

In trying to cast these limitations as mere inconvenience, opponents point out that the researcher need not be present for the duration of the screen capture for each movie.<sup>134</sup> While true, this observation does nothing to alter the ten years it would take to complete the creation of a usable corpus. DVD CCA would keep the researcher on a full-time schedule, but equip her with six computers, cutting the time to one year and four months, and perhaps add another coworker.<sup>135</sup> However, these suggestions all involve access to resources beyond the reach of nearly all researchers in relevant fields: time, financing, a phalanx of research assistants or software engineers, or a yet-to-be-invented data mining program that does not circumvent access controls.<sup>136</sup> An outright barrier dwindles to a mere inconvenience only if the researcher realistically can be expected to have the resources to overcome it. Here, that is not the case.

---

<sup>129</sup> Letter from ACH, Appendix A to Initial Comment, at 2.

<sup>130</sup> 2018 Recommendation at 78; *see also* Register of Copyrights, Section 1201 Rulemaking: Fifth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention, Recommendation of the Register of Copyrights, at 133 (Oct. 12, 2012) (“[T]here is no serious question that screen capture technology produces lower-quality images than those that are available by circumvention.”).

<sup>131</sup> Letter from David Bamman, Appendix B to Initial Comment, at 3.

<sup>132</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 6; DVD CCA & AACS LA Class 7(a) Opp’n at 14. DVD CCA does argue that screen capture must produce a sufficient digital file because David Bamman used screen capture to digitize 200 films before abandoning his project. DVD CCA & AACS LA Class 7(a) Opp’n at 14. But the fact that the project was abandoned for another of screen capture’s deficiencies does not demonstrate that screen capture is otherwise sufficient.

<sup>133</sup> Letter from David Bamman, Appendix B to Initial Comment, at 3.

<sup>134</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 6.

<sup>135</sup> DVD CCA & AACS LA Class 7(a) Opp’n at 14.

<sup>136</sup> Joint Creators Class 7(a) & 7(b) Opp’n at 6.

**d. Anti-circumvention liability is the cause of the adverse effect proponents seek to mitigate in this proceeding.**

Nothing about additional barriers imposed by contract terms changes the fact that section 1201's prohibition on circumvention causes the adverse effect proponents seek to remedy here. Opponents' ability to use licenses or browsewrap agreements to compound risks faced by researchers does not mean that the risk of section 1201 liability is somehow illusory.<sup>137</sup> First, serious questions exist as to the enforceability of some of those agreements, and the Office has properly recognized that such questions are distinct from those addressed in section 1201 proceedings.<sup>138</sup> Second, content owners are not uniformly hostile to TDM research.<sup>139</sup> Finally, the penalties for violations of section 1201 are quite different from—and more severe than—damages for breach of contract.<sup>140</sup> Petitioners have satisfied the causation requirement.

\*\*\*

TDM shows great promise for deriving new insights from collections of literary works and motion pictures. Our initial comment was supported by letters from myriad researchers and research associations specifically describing the ways in which section 1201 makes it impossible for them to pursue worthy TDM projects on in-copyrighted works, as well as a dazzling array of research projects they hope to pursue if the exemption is granted. Opponents appear to object primarily to the scope of the requested exemption, rather than to its core purpose. Petitioners do not object to accommodating legitimate concerns to the extent that the core purpose is preserved. The uses we propose to make of the exemption are noninfringing, fair uses. Opponents do not meaningfully

---

<sup>137</sup> See AAP Class 7(b) Opp'n at 10–11 (arguing that because some licenses and browsewrap agreements forbid TDM, section 1201 is not the cause of the adverse effect).

<sup>138</sup> See Recommendation of the Register of Copyrights in RM 2008-8, Rulemaking on Exemptions from Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies, at 172–73 (June 11, 2010) (“courts have recognized that contractual obligations stand apart from the copyright law. . . . This rulemaking proceeding . . . appl[ies] only to the Section 1201(a)(1) prohibition on circumvention of access controls and ha[s] no effect whatsoever on the validity or viability of such breach of contract claims.”); Kyle K. Courtney, Rachael Samberg, and Timothy Vollmer, *Big Data Gets Big Help: Law and Policy Literacies for Text Data Mining*, 81.4 Ass'n of College & Research Libraries 193 (2020), available at <https://crln.acrl.org/index.php/crlnews/article/view/24383/32222> (last visited Mar. 10, 2021) (noting that some courts require proof of awareness of browsewrap agreements).

<sup>139</sup> See Open Access Scholarly Publishing Association, *Members*, <https://oaspa.org/membership/members/> (last visited Mar. 10, 2021); Claire Redhead, *OASPA Comment on Text and Data Mining and Proposed EU Copyright Reform*, Open Access Scholarly Publishing Association (Dec. 21, 2016), <https://oaspa.org/oaspa-comment-text-data-mining-proposed-eu-copyright-reform/>.

<sup>140</sup> 17 U.S.C. § 1204(a) (applying criminal penalties to certain violations of section 1201).

contend that there are alternatives to an exemption, and indeed there are none. For these reasons, our exemption request should be granted.