UNITED STATES COPYRIGHT OFFICE

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

ITEM A. COMMENTER INFORMATION

Commenters:

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 http://www.authorsalliance.org.

Represented by:

Samuelson Law, Technology & Public Policy Clinic
University of California, Berkeley, School of Law
Erik Stallman, Associate Director, estallman@clinical.law.berkeley.edu
Jennifer M. Urban, Director of Policy Initiatives, jurban@clinical.law.berkeley.edu
Mathew Cha, Christian Howard-Sukhil, and Zhudi Huang, Clinical Law Students

(2) **The American Association of University Professors** ("**AAUP**") is a nonprofit membership association of faculty 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.

Represented by:

Risa Lieberwitz, AAUP General Counsel, rlieberwitz@aaup.org
Aaron Nisenson, AAUP Senior Counsel, anisenson@aaup.org
Edward Swidriski, AAUP Assistant Counsel, eswidriski@aaup.org

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

Represented by:

Jonathan Band
Jonathan Band PLLC
jband@policybandwidth.com

TABLE OF CONTENTS

ITEM B. PROPOSED CLASS ADDRESSED

This Reply Comment addresses Class 3(a) Motion Pictures and Class 3(b) Literary Works.

ITEM C. OVERVIEW

The proposed expansion of the current exemption for text and data mining ("**TDM**") is straightforward. Digital humanities scholars are using the current exemption to generate new insights about our culture. But as they have gained experience with the exemption, they have encountered hurdles that impose significant and unnecessary limitations on the value that the exemption can provide to the public. Specifically, the lack of a provision for sharing corpora imposes major limitations on the knowledge that can be generated because different researchers, who could bring new questions to the material and employ new methods for answering those questions, are foreclosed from doing so unless they start anew. As Professor John Bell explains, "[t]he evolving methodology of entire disciplines is being held back by the requirement to restrict a prepared corpus to its original research group."[1]

Petitioners have thus requested that the Copyright Office (the "**Office**") allow digital humanities scholars to avoid the significant, redundant, and unnecessary expenditure of time and resources needed to re-circumvent, re-clean, and re-process data in a research corpus that has already been assembled for text and data mining research conducted under the exemption.[2] That obstacle adversely affects both individual researchers and the discipline of digital humanities as a whole.

Opponents raise a number of objections:[3]

- In a proceeding intended to examine proposed expansions to current exemptions, opponents suggest that it is somehow improper to seek improvements to an exemption based on experience working with it.[4] We seek this expansion because, after roughly two-and-a-half years' experience making use of the exemption, the inability to share research corpora has emerged as the most significant obstacle to broadening the field of researchers and research

---

[1] Authors All. et al. Class 3(a) & 3(b) Initial Comment ("**Initial Comment**"), App. D: Letter from John Bell at 2.

[2] *See* Dong Nguyen et al., *How We Do Things With Words: Analyzing Text as Social and Cultural Data*, 3 Frontiers Artificial Intel., Aug. 2020, at 5, 8, https://www.frontiersin.org/articles/10.3389/frai.2020.00062/full (describing how data is cleaned, annotated, and preprocessed); digital humanities scholar Kathleen Fitzpatrick has defined the digital humanities as "a nexus of fields within which scholars use computing technologies to investigate the kinds of questions that are traditional to the humanities, or . . . ask traditional kinds of humanities-oriented questions about computing technologies" (internal citations omitted). Kathleen Fitzpatrick, *The Humanities, Done Digitally*, *in* Debates in Digital Humanities (Matthew K. Gold ed., 2012), https://dhdebates.gc.cuny.edu/read/untitled-88c11800-9446-469b-a3be-3fdb36bfbd1e/section/65e208fc-a5e6-479f-9a47-d51cd9c35e84#ch02.

[3] Opponents also include a number of objections that attack the core fair use determination that the Office has already made regarding the existing exemption. Opponents had ample opportunity to make those objections earlier in this process but failed to do so, and it is inappropriate at this stage to attempt to reopen that determination.

[4] *See* Motion Pictures Association, News/Media Alliance, and Recording Industry Association of America ("**MPA et al.**") Class 3(a) & 3(b) Opp'n at 5. *See also* Exemptions to Permit Circumvention of Access Controls on Copyrighted Works, 88 Fed. Reg. 37,486, 37,489 (proposed June 8, 2023) (to be codified at 37 C.F.R. pt. 201) (describing the process for "proposals to expand current exemptions").

projects making lawful and productive use of the exemption. It would be absurd to suggest that petitioners and the Office should ignore these experiences.

- Opponents argue that the proposed expansion would not require compliance with the existing exemption's stringent requirements.[5] This is incorrect. The research enabled by the proposed expansion would involve the same fair use of the underlying works, subject to precisely the same restrictions and conditions, as the research enabled by the current exemption. The Office has already determined that this research is fair use.[6] The requested modification would only permit an institution to share a TDM corpus with researchers at other institutions of higher education "*for the purposes of conducting independent text and data mining research and teaching, where those researchers are in compliance with this exemption.*"[7] All of the existing requirements would also apply to uses under this expansion.
- Opponents also make various arguments that seek to undermine or further restrict the current exemption instead of addressing the proposed expansion.[8] These arguments should have been raised in response to the request to renew the current exemption. In any case, nothing—including the responses to the coordinated, strategically timed mass inquiries into security measures submitted by the Association of American Publishers ("**AAP**") and the Motion Picture Association ("**MPA**")—justifies departure from the Office's stated intention to recommend renewal of the exemption.[9] If anything, the effort demonstrates why AAP and the MPA are not the proper entities to be making these inquiries.

The proposed expansion is targeted to address observed adverse effects on research and is subject to the same requirements as the current exemption. It is also vital to "ensur[ing] that *contemporary* history, culture, and society are not omitted from the scholarly record."[10] As Professor Bell stated in his letter of support: "Adding a provision to the Text and Data Mining Exemption allowing media corpora to be shared does not just make existing research easier—in many cases, it would

---

[5] *See* DVD Copy Control Association and Advanced Access Content System Licensing Administration ("**DVD CA & AACS LA**") Class 3(a) Opp'n at 13 ("Proponents incongruously argue that the proposed expansion should, in fact, allow unaffiliated researchers to use other institutions' corpora regardless of whether they too have copies of the underlying works.").

[6] Shira Perlmutter, U.S. Copyright Off., Recommendations of the Register of Copyrights, Section 1201 Rulemaking: Eighth Triennial Proceeding to Determine Exemptions to the Prohibition on Circumvention 117 (Oct. 19, 2021) [hereinafter 2021 Recommendation] ("Balancing the four fair use factors, with the limitations discussed, the Register concludes that the proposed use is likely to be a fair use"). *Authors Guild, Inc. v. HathiTrust*, 755 F. 3d 87, 97 (2nd Cir. 2014) (holding that creation of a full-text searchable database was transformative use when it did not show the user any of the text of the copyrighted works); *Authors Guild, Inc. v. Google, Inc.*, 804 F.3d 202, 217 (2d Cir. 2015) (holding that the "snippet view," which showed portions of unaltered, copyrighted text, was transformative because it "add[ed] important value to the basic transformative search function" by allowing users to verify that the list of books returned by the database was responsive to the user's search).

[7] Initial Comment at 6 (emphasis in original).

[8] *See* AAP Class 3(b) Opp'n at 16–18.

[9] Exemptions to Permit Circumvention of Access Controls on Copyrighted Works, 88 Fed. Reg. 72,013, 72,018 (proposed Oct. 19, 2023) (to be codified at 37 C.F.R. pt. 201).

[10] Initial Comment, App. M: Letter from the Mellon Foundation at 1 (emphasis in original).

make research possible that could not even be considered without it."[11] Petitioners have made the required showing to enable that research, and the Office should recommend granting this expansion of the exemption.

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

The technological protection measures ("**TPM**s") at issue in the expansion cover the same classes of works and methods of circumvention as articulated in the current exemption, namely, motion pictures on "DVDs protected by the Content Scramble System, on a Blu-ray disc protected by the Advanced Access Content System, or made available for digital download" and literary works "distributed electronically."[12]

AAP's objection that the description of these methods is inadequate is inapposite. Digital humanities researchers work with the media formats and protections chosen by copyright holders over time—accordingly, the precise format and TPM version may vary according to copyright holders' choices. The TPMs in question and the methods used for circumventing them are and will be those necessary for these formats. Further, these are the same measures used in the current exemption; we are requesting no changes for our proposed expansion.[13] AAP's objection is thus also beyond the scope of this proceeding. As to who is performing the circumvention, this is also covered by the current exemption.[14] The remainder of AAP's discussion of the issue—and a large part of the opposition from the International Association of Scientific, Technical and Medical Publishers ("**STM**")—addresses descriptions of pirate sites and commercial AI activity.[15] None of this is relevant to digital humanities researchers' circumvention of TPMs in order to engage in noninfringing TDM research.

ITEM E. ASSERTED ADVERSE EFFECTS ON NONINFRINGING USES

The value of the current exemption for TDM is clear: it allows digital humanities researchers to embark on exciting and socially valuable scholarship. At the same time, experience with the exemption has shown that the limitation on sharing has significant adverse effects on non-infringing TDM research and teaching. The requested expansion is needed to address these effects. Opponents' various attempts to argue otherwise are unavailing; we respond to each in turn.

---

[11] Initial Comment, App. D: Letter from John Bell at 2.

[12] 37 C.F.R. § 201.40(b)(4)(i) & (b)(5)(i).

[13] Initial Comment at 19 ("The proposed expansion does not materially alter the nature and basic operations of the relevant technological protection measures and methods of their circumvention as compared to the current exemption.").

[14] Petitioners direct AAP to 37 C.F.R. § 201.40(b)(5)(i)(A) ("The circumvention is undertaken by a researcher affiliated with a nonprofit institution of higher education, or by a student or information technology staff member of the institution at the direction of such researcher.").

[15] AAP Class 3(b) Opp'n at 5–6; STM Class 3(b) Opp'n at 2.

1. **The inability to share corpora has meaningful, debilitating, adverse effects on researchers' ability to engage in socially valuable research.**

The need for the proposed expansion is well-documented in the Initial Comment and the supporting letters from researchers, legal scholars, and librarians.[16] Contrary to AAP's unsupported statements otherwise, these adverse effects on valuable, noninfringing research and teaching go far beyond "*de minimis* impacts" and are clearly "distinct, verifiable and measurable."[17]

First, the costs associated with the prohibition on sharing TDM corpora under the current exemption are significant, pervasive, and debilitating. Letters supporting the petition explain this in detail. For instance, some supporting letters described "laborious" digitization processes[18] that can require "massive redundancy" to make a corpus research-ready;[19] others discussed the sheer costs of creating a corpus, which can amount to "tens of thousands of dollars" independent of acquisition;[20] still others relayed that they spent so much time putting their corpus together that at least one researcher has required "an unfunded extension to complete the study itself."[21] As also established in the Initial Comment, these costs prevent scholars from conducting valuable research and can bar less well-resourced scholars from contributing to the field altogether.[22] Accordingly, precluding researchers from working with existing corpora to ask new questions and build on previous research imposes clear adverse effects on researchers' ability to engage in socially valuable, noninfringing research.[23]

Opponents Motion Pictures Association, News/Media Alliance, and Recording Industry Association of America ("**MPA et al.**") and AAP also make much of the fact that we proposed "collaboration" and "replication" as allowable practices in the 2021 Proceeding.[24] The current exemption certainly can support some valuable collaborative work. But experience since the last

---

[16] *See*, *e.g.*, Initial Comment, App. B: Letter from Mark Algee-Hewitt at 2, App. E: Letter from Joel Burges and Emily Sherwood at 2, App. K: Letter from Lauren Tilton and Taylor Arnold at 2.

[17] AAP Class 3(b) Opp'n at 6 (quoting 2021 Recommendation at 12 (quoting H.R. Rep. No. 105-551, pt. 2, at 37 (1998)).

[18] Initial Comment at 12 (quoting App. C: Letter from David Bamman at 2).

[19] *Id*. (quoting App. E: Letter from Joel Burges and Emily Sherwood at 2).

[20] *Id*. (quoting App. L: Letter from Henry Alexander Wermer-Colan at 2). To be clear, thousands of dollars in digital humanities research, where grants are both scarce and small compared to some other disciplines, can represent an insurmountable sum for some researchers, especially if it is necessary to spend this sum anew with each additional researcher. *Id.* at 13 (referencing App. J: Letter from Rachael Samberg and Timothy Vollmer at 3).

[21] Initial Comment, App. D: Letter from John Bell at 1–2.

[22] Initial Comment at 14–15 (citing App. B: Letter from Mark Algee-Hewitt at 3); Initial Comment at 8–10.

[23] Digital humanities TDM research relies on specialized methods to prepare and analyze a research corpus. Accordingly, we have provided a more detailed example of data preparation used for one TDM research project in Appendix A to this Reply Comment to illustrate some example methods and provide a sense of the time and effort required.

[24] MPA et al. Class 3(a) & 3(b) Opp'n at 3–6; AAP Class 3(b) Opp'n at 7.

proceeding has shown that the allowance for collaboration is not sufficient to support a wide swathe of socially beneficial digital humanities research.

First, the allowance for collaboration cannot support new, independent work that builds on an existing corpus. Researchers' ability to independently raise questions about a corpus is just as crucial to the progress of digital humanities research as collaboration and replication. Enabling new researchers to ask new questions in the field "catalyzes the speed and quality of TDM research."[25]

Relatedly, research validation requires methods that are more flexible than "solely for purposes of . . . replication of the research."[26] Replication typically involves testing whether the same research method would lead to the same results. In comparison, validation would include using different research methods to test out a particular research result.[27] For the latter, the ability to use the same corpus is essential. Validations using different research methods employed by new researchers fall outside the current exemption's limitation on corpora sharing "solely for purposes of . . . replication of the research," but are just as critical for robust studies and the development of knowledge.

Second, opponents' argument that the term "collaboration" is not ambiguous[28] unfortunately lacks grounding in experience with TDM research. As explained fully in the Initial Comment, on-the-ground experience shows that researchers find the exact scope and nature of "collaboration" under the current exemption unclear.[29] This confusion prevents this provision from supporting TDM research as originally intended.[30] For instance, researchers, who "tend to be conservative in their interpretation of what is allowed,"[31] are unsure of when a connection is sufficient to qualify as a "collaboration." This comes up in a variety of contexts—from questions about the stage or formality of the potential collaboration[32] to questions about what happens when a research team member working on a relatively discrete project moves institutions,[33] and even extends to questions about whether students qualify as collaborators.[34] This confusion clearly asserts adverse effects on noninfringing TDM research, ultimately to the detriment of the public.

---

[25] Initial Comment at 16–19 (quoting App. M: Letter from the Mellon Foundation at 1).

[26] 37 C.F.R. § 201.40(b)(4)(i)(D) & (b)(5)(i)(D).

[27] *See* Initial Comment, App. D: Letter from John Bell at 2 ("[T]he first thing our readers will want to do is rerun the analysis using new models to produce more accurate results or examine a related research question that could not be addressed using current inference models.").

[28] AAP Class 3(b) Opp'n at 7.

[29] Initial Comment at 8–10.

[30] *Id*.

[31] *Id*. at 9.

[32] *Id*. at 8–9 (quoting App. B: Letter from Mark Algee-Hewitt at 2; App. M: Letter from the Mellon Foundation at 2).

[33] *Id*. at 9–10 (quoting App. B: Letter from Mark Algee-Hewitt at 2).

[34] *Id*. at 9 (citing App. E: Letter from Joel Burges and Emily Sherwood at 2).

Granting the proposed expansion will ameliorate these adverse effects. The proposed expansion is targeted to address these identified adverse effects and does not disturb any other existing requirements.[35] Rather, it is circumscribed by the exact same guardrails as the current exemption.[36] In light of these built-in guardrails, this proposed expansion is a limited, modest step to ensure that the current exemption can practically be utilized by academic researchers.

2. The Office's fair use analysis of the current exemption applies equally here and leads to the same conclusion of a noninfringing fair use.

In its 2021 Recommendation, the Office concluded that text and data mining for scholarly research and teaching was likely to be a non-infringing, transformative fair use because the function and purpose of the use differed from the original works.[37] Citing no actual changes in controlling precedent, opponents nonetheless devote most of their fair use argument to attacking the current exemption. Those arguments rest largely on features of a research project that rely on a separate § 1201 exemption and on concerns about commercial artificial intelligence applications that are irrelevant here. On the separate issue of security measures, opponents rely on letters of inquiry into security measures and responses to these inquiries. These inquiries, however, chiefly demonstrate the wisdom of the Office's designation of *copyright owners* as the appropriate party to make those inquiries, rather than trade associations. None of these arguments undermines the fair use case for the current exemption or proposed expansion.

a. The purpose and character of the proposed use remains noncommercial, nonexpressive, and transformative.

No intervening factual or legal developments disturb the Office's previous conclusion that the use of works under the current exemption is for a transformative, noncommercial, and statutorily favored purpose.[38] The proposed expansion concerns precisely the same purpose. For the most part, opponents focus their arguments regarding the first factor on the current exemption rather than the proposed expansion. These arguments would have been appropriate in response to the

[35] Contrast with MPA et al. Class 3(a) & 3(b) Opp'n at 5 ("Their proposed new language would dramatically enlarge the scope of the exemption adopted in 2021—which the proponents proposed in their 2021 reply comments—although it isn't clear how far they would like it to extend or how it would possibly work in practice.").

[36] The requirements are: (a) The group of users is limited to researchers, eligible employees, or students affiliated with nonprofit higher-education institutions; (b) The exemption is for the sole purpose of conducting TDM research and teaching; (c) Such institutions are required to adopt effective security measures as defined under 37 C.F.R. § 201.40(b)(4)(ii)(B) & (b)(5)(ii)(B); (d) Such institutions must own lawful copies of the underlying works or else license these works without a time limitation on access. 37 C.F.R. § 201.40(b)(4) & (5).

[37] 2021 Recommendation at 109.

[38] *See id*. (holding that "a use can be transformative if the function or purpose of the use differs from that of the original" and that providing information about the works is different from the "expressive or informative purposes of the original works.").

petition for renewal; they are not appropriate here.[39] Regardless, none of the arguments is an adequate basis for the Office to reverse its prior conclusion regarding the purpose and character of TDM research at issue here.

First, opponents—and DVD Copy Control Association and Advanced Access Content System Licensing Administration ("**DVD CA & AACS LA**") in particular—make much of the "close viewing" that Kinolab enables to argue that researchers are using the current exemption for nontransformative purposes.[40] But as clearly explained in Allison Cooper's letter describing Kinolab, the close-viewing aspects of the lab's work do not rely on the exemption at issue here.[41] Instead, those aspects rely on the exemption for making short portions of motion pictures for "criticism, comment, teaching, or scholarship."[42] Indeed, Kinolab has been performing this work since at least 2019, years before the TDM exemption existed.[43] Accordingly, opponents' arguments about Kinolab's "close viewing" are outside the scope of this proceeding.

Relatedly, DVD CCA & AACS LA's comparison between Kinolab's annotation project and the lexicon found infringing in *Warner Bros. Entertainment Inc. v. RDR Books*, 575 F. Supp. 2d 513 (S.D.N.Y. 2008) and similar cases is both beside the point and incorrect.[44] Again, the "close viewing" method that Kinolab employs relies on a separate exemption. Further, the annotation and conversion of film clips go to a different purpose that the Lexicon in *RDR* Books: a table of descriptive metadata or a book analyzing the evolution of the close-up is not a substitute for even a short portion of a film.[45]

DVD CCA & AACS LA's concerns about close viewing and annotation must also be distinguished from the work necessary for "converting the materials into a format compatible with computational

---

[39] Exemptions to Permit Circumvention of Access Controls on Copyrighted Works, 88 Fed. Reg. at 37,487 ("If the Office recommends renewal of the current exemption, then it will consider *only* the discrete aspects relevant to its expansion as a new petition.") (emphasis added).

[40] DVD CCA & AACS LA Class 3(a) Opp'n at 3–5. We note that DVD CCS & AACS LA appear to miss the fact that "close viewing" is a term specific to film studies and can refer to a range of techniques that are distinct from the extracted semantic metadata of images associated with "distant viewing." *See* Taylor Arnold and Lauren Tilton, *Distant Viewing: Analyzing Large Visual Corpora*, 34 Digit. Scholarship Humans. i3, i3 (2019), https://doi.org/10.1093/llc/fqz013.

[41] *See* Initial Comment, App. G: Allison Cooper at 3 ("Our work could not legally take place without the exemptions permitting the circumvention of TPM for the purpose of criticism, comment, teaching, or scholarship and, more recently, the exemption permitting limited cross-institution collaborations for the purpose of TDM.").

[42] 37 C.F.R. § 201.40(b)(1)(ii)(A).

[43] *See* Tom Porter, *Bowdoin Selected for Pioneering Computer Science Ethics Challenge*, Bowdoin (Apr. 30, 2019), https://perma.cc/LR7B-XZMJ.

[44] DVD CCA & AACS LA Class 3(a) Opp'n at 3–4. Nor do TDM researchers engage in "space-shifting" as described in the caselaw; opponents' comparisons to these cases are inapposite. MPA et al. Class 3(a) & 3(b) Opp'n at 8.

[45] *RDR Books*, 575 F. Supp. 2d at 535. This distinction is abundantly clear if one views the content of the annotation tags that DVD CCA & AACS LA has cropped from the clip of *In the Mood for Love* that they included in their appendix. *See* DVD CCS & AACS LA Class 3(a) Opp'n, Ex. 9–11. An uncropped version is included in the documentary evidence. App. B: Uncropped screenshot of Kinolab.

analysis (e.g. plain-text computer files for literature, video files for motion pictures)."[46] That work may involve deleting text or marking text to distinguish bibliographic information from the content of a literary work. That conversion and processing is necessary for the "machine generation of metadata [that] is the only realistic way to examine these collections at scale."[47] The work may also involve labeling materials with metadata for use in computational analysis.[48] Either way, this processing is part of the transformative and nonexpressive use embraced by the exemption and the fair use doctrine.

Second, opponents point to brief references to artificial intelligence in supporting letters, arguing that researchers are using the exemption for impermissible purposes.[49] Opponents made the same arguments in the last § 1201 proceeding.[50] It is again unclear why opponents did not raise this argument as an opposition to renewal, as it goes to the current exemption rather than the proposed expansion.

In any event, the argument should fail. The evolution of research methods that may include machine learning does not undermine their legitimacy.[51] Distant viewing, for example, "emphasizes the at-scale computational analysis of digital images through machine learning."[52] AAP's proposal to declare such methods off-limits and confine TDM to a crabbed understanding of "statistical methods" is both untimely and misguided.[53] An article discussing the computational analysis of conversations in film is not a substitute for watching a movie regardless of whether a fixed algorithm or a neural network performed the underlying analysis. The uses contemplated by the exemption and conducted pursuant to it are transformative.

Noncommercial Use. The purpose of the proposed expansion remains noncommercial, and the terms of the current and proposed exemption expressly limit the proposed expansion to scholarship

---

[46] *See* Initial Comment, App. A: Letter from the Association of Computers and the Humanities at 1.

[47] Initial Comment, App. D: Letter from John Bell at 1.

[48] *See* Bakels et al., *Matching Computational Analysis and Human Experience: Performative Arts and the Digital Humanities*, 14 Digit. Humans. Q., Apr. 2020, at ¶ 44, https://perma.cc/TV7S-MM4P; Matthew Sims & David Bamman, *Measuring Information Propagation in Literary Social Networks*, *in* Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) (Bonnie Weber et al. ed., 2020), https://aclanthology.org/2020.emnlp-main.47/. *See also* App. A: Example of Dataset Preparation and Computational Analysis.

[49] *See* AAP Class 3(b) Opp'n at 12.

[50] *See, e.g.*, 2021 DVD CCA & AACS LA Class 7(a) Opp'n at 7–8 (quoting and taking issue with references to "AI, machine learning, and audiovisual data" in a supporting letter); 2021 AAP Class 7(b) Opp'n at 5 ("A collection of human-authored works can be extremely valuable to users who seek to develop artificial intelligence ('AI') and machine learning capabilities and applications, including academic researchers seeking to partner with commercial entities and/or publicize or seek remuneration for their efforts in this area.").

[51] *See* Letter from Evelyn L. Remaley, Acting Assistant Sec'y for Commc'ns & Info. & Adm'r, Nat'l Telecomms. & Info. Admin., U.S. Dep't of Commerce, to Shira Perlmutter, Register of Copyrights and Dir., U.S. Copyright Office 56 n.288 (Oct. 1, 2021).

[52] Initial Comment, App. E: Letter from Joel Burges and Emily Sherwood at 1.

[53] AAP Class 3(b) Opp'n at 17–18.

and teaching at nonprofit institutions.[54] AAP's invocation of "AI data laundering" is unrelated to anything researchers are actually doing with the exemption or hope to do with the proposed expansion.

Scholarship and Teaching. The eligible uses for both the current exemption and the proposed expansion are limited to the statutorily favored purposes of scholarship and teaching.[55] Invoking Kinolab, MPA et al. raise concerns that the exemption will be used for "showing an entire motion picture in a class" and therefore should be limited to "teaching about TDM."[56] The methods and insights of research and teaching enabled by the exemption are as suited to a course about contemporary literature as to a course about data science. However, in neither case is the exemption being used for classroom display of short portions of motion pictures, let alone films in their entirety.

Because the use of the underlying works is identical under the current exemption and proposed expansion, the Office's previous determination on the first factor applies with equal force here.

b. The analysis of the second and third fair use factors remains unchanged from that in the Office's 2021 Recommendation.

Although the proposed expansion includes literary works and motion pictures, this factor is of "limited significance" given the nature of the use at issue.[57]

Similarly, as the Office has recognized, copying a substantial amount of works to obtain data about the copyrighted work is reasonable in light of the legitimate purpose of conducting TDM research and teaching.[58] Since the proposed expansion does not change the legitimate purpose that the Office has recommended for renewal, the amount and substantiality of the copying is still reasonable and weighs in favor of fair use.

c. The proposed expansion is consistent with the Office's analysis of the fourth factor in the 2021 Recommendation.

The Office's analysis and recommended solutions for concerns related to market harm apply here. Digital humanities scholarship is not a market substitute for the works studied. There is still no commercially available licensing market suitable for digital humanities scholarship on contemporary cultural history, nor would such a market be within copyright owners' exclusive right to control. Finally, opponents' dissatisfaction with the responses that trade associations

[54] 37 C.F.R. § 201.40(b)(4)(i) & (5)(i); Initial Comment at 5–6.

[55] 37 C.F.R. § 201.40(b)(4)(i)(A) & (b)(5)(i)(A); 17 U.S.C. § 107. Accordingly, developing a generative AI platform or translation service for commercial purposes would thus fall outside the exemption.

[56] MPA et al. Class 3(a) & 3(b) Opp'n at 7–8.

[57] 2021 Recommendation at 111.

[58] *Id.*; *see also* Initial Comment, App. I: Letter from Matthew Sag at 2 (noting that transformative and nonexpressive use of a significant amount of copyrighted works is still fair use because such use does "not interfere with the interest in original expression that copyright is designed to protect.").

received to their coordinated inquiry into security practices does not warrant opponents' desired suspension or amendment of the current exemption. We address these points in turn.

i. The proposed expansion would not result in market substitutes for the original works.

Opponents spend significant time recycling arguments from the 2021 Proceeding that TDM research would affect the market for copyrighted works. In its 2021 Recommendation, the Office concluded that "as in *HathiTrust*, the end goal of the contemplated TDM research does not serve as a substitute for the original work."[59] That conclusion still holds. An article about how information propagates through character dialogue in novels is no substitute for the novel itself.[60] Likewise, a video-based corpus that has been cleaned and coded for computational analysis cannot substitute for a film. Nothing about the proposed exemption alters these conclusions.

Nor does the proposed expansion remove or relax the requirement that the receiving institution lawfully acquire a copy of each work contained in a given corpus. Because the receiving institution must lawfully acquire its own copy of the work, there is no change in the market between an institution purchasing the work for the purpose of circumventing TPMs under the current exemption and an institution purchasing the work to qualify as a receiving institution under the proposed expansion.[61] Opponents ask how such a requirement would be enforced.[62] The answer is that it is common practice for researchers to document the contents of their corpus, including "how the works were assembled, reasons the works were chosen, and why the choices were made."[63]

ii. The proposed expansion would not harm cognizable licensing markets.

AAP points to licensing arrangements that rightsholders have struck with OpenAI and Bandlab as evidence of harm to their licensing markets.[64] Yet neither a copyright owner's eagerness to create a licensing market nor a user's willingness to seek a license for a particular use is sufficient to establish cognizable market harm when the underlying use is transformative.[65] And the use at issue here is highly transformative.

Further, none of these commercial offerings are suitable for the digital humanities projects that researchers wish to undertake. AAP points to licensing platforms like the Copyright Clearance

---

[59] 2021 Recommendation at 113.

[60] *See* App. A: Example of Dataset Preparation and Computational Analysis.

[61] *Cf. Authors Guild, Inc. v. Google, Inc.*, 804 F.3d at 229 ("In these circumstances, Google's creation for each library of a digital copy of that library's already owned book in order to permit that library to make fair use through provision of digital searches is not an infringement.").

[62] MPA et al. Class 3(a) & 3(b) Opp'n at 8.

[63] Initial Comment, App. B: Letter from Mark Algee-Hewitt at 3.

[64] *See* AAP Class 3(b) Opp'n at 14.

[65] *See Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 614 (2d Cir. 2006) ("Appellant asserts that it established a market for licensing its images, and in this case expressed a willingness to license images to DK. Neither of these arguments shows impairment to a traditional, as opposed to a transformative market.").

Center's RightFind as an example of both a market harmed by the exemption and an alternative to circumvention.[66] But AAP offers no more evidence to show that RightFind is a serviceable alternative to circumvention than it did during the last triennial proceeding. Nothing has changed that would cause the Office to alter its conclusion that hitherto digital libraries are "insufficient to allow researchers to conduct the type of research contemplated here."[67]

iii. Researchers are abiding by the exemption's security requirements, which are sufficient to eliminate concerns over uncontrolled dissemination.

Opponents assume that circumventing TPMs to build a TDM corpus means that the corpus is then "decrypted" and would remain decrypted when shared.[68] That is incorrect. For example, at the University of California, Berkeley, the minimum security standards for highly confidential information[69] require sending the data using "industry-accepted encryption technologies."[70] The networked devices themselves are highly secured, and institutional devices such as servers use isolated networks with intrusion detection, logging, and firewalls with "the most restrictive rules possible."[71] Similarly, Bowdoin College maintains security covered by their Written Information Security Program, which requires that "the transmission of all sensitive data will be encrypted in transit using secure encryption algorithms and methodologies."[72] The network itself must also undergo annual internal and external security audits using the National Institute of Science and Technology's Cybersecurity Framework.[73] Accordingly, opponents' concerns on this front are unfounded. The standards, in addition to specifying how corpora are protected while at the institution, will apply to both data-in-transit and data-at-rest.

Opponents' fears that the sharing itself presents a security risk due to lax security at the receiving institution are likewise unfounded.[74] The current exemption and the proposed expansion are subject to identical, significant security requirements. Indeed, these requirements are unique among the temporary exemptions to liability under § 1201 in that they require measures equivalent

[66] AAP Class 3(b) Opp'n at 14.

[67] 2021 Recommendation at 117.

[68] AAP Class 3(b) Opp'n at 2, 3, 7, 16, 18.

[69] *Data Classification Standard*, Berkeley Info. Sec. Off., https://perma.cc/D8M7-DH36.

[70] Minimum Security Standards for Electronic Information v3 § 6.1 Encryption in Transit (Feb. 29, 2024), https://perma.cc/6WNB-PSZL (available via https://security.berkeley.edu/minimum-security-standards-electronic-information).

[71] *Id*. at § 11 – Network Security.

[72] App. C: Response from Bowdoin College to the MPA at 3.

[73] *Id.*

[74] *See* AAP Class 3(b) Opp'n at 16; MPA et al. Class 3(a) & 3(b) Opp'n at 9–10; STM Class 3(b) Opp'n at 2.

to those the institution uses for its own highly confidential information.[75] Some researchers who hope to make use of shared TDM corpora have specified that they have used, or worked to design themselves, customized and protected storage environments.[76] Others have specified that they use secure encryption practices at all times.[77] And still others have specified that when using high performance computing environments, they use protected storage designed to hold "protected health information," "other highly sensitive human subjects data," and "controlled unclassified information."[78] These are precisely the measures these institutions use for their own highly confidential information and these measures precisely meet the requirements of the exemption.

iv. The trade associations' letters of inquiry were ill-targeted and overbroad, and the responses to them were appropriate.

There were significant irregularities with the senders, recipients, and timing of the letters that AAP and the MPA sent to nearly every individual who wrote a letter in support of the proposed expansion. While the 2021 Recommendation contemplated that inquiries into security measures could factor into future triennial proceedings,[79] the trade association effort here is not in line with that recommendation.

First, the exemption clearly conditions the exemption on furnishing security information to a "*copyright owner*" upon request.[80] AAP's proposed amendment to the contrary is both untimely and unsound.[81] While the trade associations purport to represent their members in making these inquiries, there are reasons to question that assertion. In particular, it is unlikely that some of the publishers AAP lists in its letter—for example, university presses associated with institutions where targeted researchers work—were even aware that this inquiry was being made. It is also

---

[75] By comparison, other § 1201 exemptions generally require users to employ reasonable security measures. *See, e.g.*, 37 C.F.R. § 201.40(b)(1)(ii)(B) (requiring nonprofit education institutions offering massive open online courses to "appl[y] technological measures that *reasonably* prevent unauthorized further dissemination"), § 201.40(b)(2)(i)(C) (requiring educational institutions providing accessible media to store such media "in a manner intended to *reasonably* prevent unauthorized further dissemination of a work") (emphasis added), § 201.40(b)(3)(ii)(E) ("The library, archives, or museum implements *reasonable* digital security measures as appropriate for the activities permitted by paragraph (b)(3)(i) of this section.") (emphasis added), § 201.40(b)(6–15) (listing no security requirements), § 201.40(b)(17) (stating that libraries, archives, or museums engaged in video game archival work "implement[] *reasonable* digital security measures as appropriate for the activities permitted") (emphasis added), § 201.40(b)(18) (holding that libraries, archives, or museums engaged in computer program archival work "implement[] *reasonable* digital security measures as appropriate for the activities permitted") (emphasis added).

[76] *See* UC Berkeley's Response to AAP's Letter (AAP Class 3(b) Opp'n, Ex. 3 at 26); *see also Secure Research Data & Computing*, Berkeley Rsch. I.T., https://perma.cc/8PLE-6NWE; Initial Comment, App. D: Letter from John Bell at 1.

[77] App. C: Response from Bowdoin College to the MPA.

[78] *Secure Research Data & Computing*, *supra* note 76.

[79] 2021 Recommendation at 117.

[80] 37 C.F.R. § 201.40(b)(4)(ii)(B) & (b)(5)(ii)(B) (emphasis added).

[81] AAP Class 3(b) Opp'n at 17.

unlikely that those publishers would agree with AAP that their associated institutions should no longer be able to conduct TDM research under the exemption.

Second, the trade associations did not attempt to accurately target their requests. The MPA sent letters to researchers who were clearly engaged in studying textual works, and AAP sent letters to researchers who were clearly engaged in studying audiovisual works. They even sent letters to individuals who supported the proposed expansion but whose support letters show they do not actually make use of the exemption. For example, Professor Matthew Sag's letter of support begins by noting that he is a Professor of Law in Artificial Intelligence, Machine Learning, and Data Science, as well as "an expert on the legal issues relating to TDM research, particularly in relation to copyright law."[82] His letter of support makes no reference to personally using copyrighted works for TDM purposes and provides no indication that Professor Sag is using the exemption. The response from Professor Sag and his campus counsel—pointing out that the trade association was mistaken in directing a letter to him—was entirely appropriate. Similarly, Brandon Butler makes clear in his letter of support that he is not a TDM researcher and that his letter is based on an article that he co-authored with Pat Aufderheide and Kimberly Anastacio, in which the authors "conducted in-depth interviews with TDM researchers" from around the world.[83] Thus, the University of Virginia counsel was understandably "at something of a loss as why his comments would cause AAP to draw any conclusion about the activities at the University of Virginia."[84]

The AAP and MPA letters make clear that participation in this proceeding was the sole criterion used to determine the scope of its inquiry.[85] Copyright owners should make some effort to investigate whether a researcher or institution is actually making use of the exemption before sending a demand letter with the implication that a response is required under the cited regulation. It is also notable that, according to the information provided by the trade associations, it appears that no researcher at an institution who received Mellon Foundation funding under the TDM grant but did *not* write a letter in support of the proposed expansion received a letter of inquiry.[86]

The timing of letters also casts doubt as to the substantive value of these inquiries. AAP and the MPA sent out letters roughly a week apart in late January and early February, giving recipients only a few weeks to respond. That timing worked well for the trade associations, which could make the most of either a rushed response or no response by the arbitrary deadline given in the letters. This timing, however, did not work well for researchers who, consistent with their institutions' practices, had to flag the inquiries for campus counsel. Despite AAP's assertion to the contrary, the only conclusion to be drawn from institutions who said they needed more time was that they needed more time.[87]

---

[82] Initial Comment, App. I: Letter from Matthew Sag at 1.

[83] Initial Comment, App. F: Letter from Brandon Butler at 1.

[84] AAP Class 3(b) Opp'n, Ex. 3 at 5.

[85] *See, e.g.*, AAP Class 3(b) Opp'n, Ex. 2 at 1.

[86] The University of Cincinnati also received a Mellon Foundation grant, AAP Class 3(b) Opp'n, Ex. 4 at 4, but AAP provides no indication of sending them a letter. *Id.* Ex. 1 at 20 (providing a table listing letter recipients).

[87] *See* AAP Class 3(b) Opp'n at 10 n.5 (discussing responses and requests for more time).

Finally, the substantive responses that the trade associations received were appropriate. In response, institutions like Bowdoin provided copies of their security and data classification policies and explained that these were communicated to TDM researchers.[88] The response, sent by Bowdoin's legal counsel, also invited further inquiries.[89] That response was entirely consistent with the exemption's requirement that "[i]f the institution uses the security measures it uses to protect its own highly confidential information, it must, upon a reasonable request from a copyright owner whose work is contained in the corpus, provide information to that copyright owner regarding the nature of such measures."[90]

3. The proposed expansion is supported under each of the five statutory factors.

The proposed expansion would enable more researchers to conduct ambitious projects to grow the field of digital humanities without harming or undermining the availability or market for the original works. The proposed expansion would also enhance the diversity of researchers, institutions, and projects contributing to that field. Accordingly, the § 1201 statutory factors favor the expanded exemption. Opponents make few attempts to argue otherwise,[91] but we respond to their stated concerns below.

a. The proposed expansion would not diminish the availability for use of copyrighted works.

With respect to the first and second § 1201 statutory factors, AAP's stated concern that the proposed expansion would "discourage the dissemination and availability" of works[92] is unfounded. As to the first factor,[93] the proposed expansion would not decrease the availability for use of works. The Office correctly found last cycle that "[t]he proposed use is narrowly tailored to scholarly research, and it is unlikely that copyright owners would entirely withhold electronic versions of their works from the market out of a concern that they may be used for the type of research described here."[94] This remains true since the proposed expansion contemplates the same types of noninfringing uses governed by the same restrictions as the current exemption. Contrary to AAP's worries of works being "exposed to piracy,"[95] security risks are minimal because, as

[88] App. C: Response from Bowdoin College to the MPA.

[89] App. C: Response from Bowdoin College to the MPA.

[90] 37 C.F.R. § 201.40(b)(4)(ii)(B).

[91] MPA et al. simply make the conclusory statement that "proponents fail to establish . . . that the Section 1201(a)(1)(C) factors, as properly construed, support granting the proposed exemption." MPA et al. Class 3(a) & 3(b) Opp'n at 5–6. DVD CCA & AACS LA do not mention the § 1201 statutory factors; nor does STM. AAP does not engage in detail with the factors, but provides some brief discussion in its comment. AAP Class 3(b) Opp'n at 15–16. We take its arguments in turn.

[92] AAP Class 3(b) Opp'n at 16.

[93] 17 U.S.C. § 1201(a)(1)(C)(i).

[94] 2021 Recommendation at 119–20.

[95] AAP Class 3(b) Opp'n at 16.

described more fully above, the sharing contemplated under this exemption would only occur between academic institutions under secure conditions.

Further, by making TDM corpora more accessible to other researchers and their institutions, opponents would likely benefit because institutions would need to lawfully acquire the works necessary to receive a given corpus.[96] In short, the expansion may actually enhance the market for the copyrighted works and thus increase the use of works for scholarly and educational purposes.

b. The proposed expansion would increase the availability for use of works for nonprofit archival, preservation, and educational purposes by lowering barriers to TDM research.

With respect to the second § 1201 statutory factor, opponents do not directly rebut the fact the proposed expansion would enhance "the availability for use of works for nonprofit archival, preservation, and educational purposes."[97] Enhancing the availability of works for TDM research and teaching is a fundamental benefit of the proposed expansion.

Instead, opponents attack funding efforts to support that research.[98] The Mellon Foundation's funding, while key to enabling the socially beneficial TDM research covered by the current exemption, also highlights the very need for the expansion. As AAP points out, $1 million of grant funding has been enough to support only *six* TDM projects.[99] While the fact that a grant-funding institution can support such projects (as a result of the initial 2021 TDM exemption) is progress, the time, effort, and funding that it takes to transform a set of digital works into a corpus ready for analysis still present significant barriers to TDM research. In other words, academic scholars who have not received external grant funding have largely been unable to conduct TDM research because of the additional costs imposed by the exemption's restrictions. Indeed, one of the researchers, despite receiving a grant, had to request unfunded extensions just to complete the process of preparing a set of copyrighted works for text and data mining.[100] The expansion of the TDM exemption to allow corpora sharing among TDM scholars would lower this significant hurdle.

To be clear, AAP incorrectly characterizes the Mellon Foundation's purpose in issuing the grant when it says that the grant was "to induce the Copyright Office to expand the TDM exemption" to allow sharing.[101] The Mellon Foundation has funded text and data mining projects as far back as

[96] 37 C.F.R. §§ 201.40(b)(4)(i)(B) & 201.40(b)(5)(i)(B).

[97] 17 U.S.C. § 1201(a)(1)(C)(ii).

[98] AAP Class 3(b) Opp'n at 3.

[99] *Id*.

[100] Initial Comment at 12 (citing App. D: Letter from John Bell at 1–2).

[101] AAP Class 3(b) Opp'n at 3.

2006,[102] long before the TDM exemption was even contemplated. It did so in furtherance of its mission "to build just communities enriched by meaning and empowered by critical thinking."[103] Those projects, however, used exclusively public-domain works. The frailties, inaccuracies, and limitations of digital humanities TDM projects relying entirely on public domain works are well-known.[104] Now that at least some researchers are able to apply this important research method to in-copyright works, providing researchers with the resources to pursue these projects is entirely within the Mellon Foundation's mission. To be sure, one goal of the Mellon Foundation's grant is to demonstrate that the *current exemption* for TDM is being used and should be renewed (not to demonstrate need for an expansion). The Mellon Foundation's support for digital humanities research making use of the temporary exemption is entirely proper and furthers the purposes of the Copyright Act and the second statutory factor.

In its 2021 Recommendation, the Office noted that issues of equity and diversity relate to the second and third § 1201 factors.[105] The proposed expansion would increase equity in research—and thus the quality and quantity of the research—by lowering barriers for researchers at smaller institutions.[106] Opponents make no argument to the contrary.

c. The proposed expansion would further the statutorily favored purposes of scholarship, research, and teaching.

The third § 1201 statutory factor considers "the impact that the prohibition on the circumvention of technological measures applied to copyrighted works has on criticism, comment, news reporting, teaching, scholarship, or research."[107] With respect to this factor, AAP refers back to its argument that the current exemption's allowance for collaboration is sufficient.[108] As we explain in Item E.1 above, this assertion is incorrect. Unfortunately, the current exemption's limitation on corpora sharing significantly limits the freedom of researchers who are unable to bear the additional costs of compiling a research-ready corpus even after they have procured copies or unrestricted licenses to the underlying works. Academic research requires a full exchange of ideas among researchers—particularly those affiliated with different institutions—to reduce bias and ensure research sustainability. This exchange is currently compromised by the existing exemption's limitations. The proposed expansion, on the other hand, furthers the purposes covered under this factor. It

[102] *Developing Text-Mining in the Digital Library – MONK*, Mellon Found., https://perma.cc/XP57-MCMF. The description accompanying the grant states that the grant is intended "to support the development of a system for the mining and analysis of large-scale corpuses of digitized text."

[103] *Mission*, Mellon Found., https://perma.cc/3V93-WCV5.

[104] Authors All., Am. Ass'n of Univ. Professors, & Libr. Copyright All., Round 1 Comment on Notice of Proposed Rulemaking on Exemptions To Permit Circumvention of Access Controls on Copyrighted Works 21 (Dec. 13, 2020), https://perma.cc/RZK6-WRLD.

[105] 2021 Recommendation at 120 n.667.

[106] Initial Comment at 33 ("[T]he ability to conduct research using [the current] exemption depends largely on the resources available at each institution").

[107] 17 U.S.C. § 1201(a)(1)(C)(iii).

[108] AAP Class 3(b) Opp'n at 16.

would enable more—and more rigorous—scholarship and research. Ultimately, it would "create a more efficient research pipeline and speed up discovery and the advancement of knowledge."[109] Granting the proposed expansion would provide TDM researchers with the much-needed ability to engage in the kind of scholarship that the Office recognized as being academically and socially beneficial in the 2021 Proceeding.[110]

d. The proposed expansion is not likely to harm the market for or value of copyrighted motion pictures and literary works used in TDM research.

The fourth § 1201 statutory factor considers "the effect of circumvention of technological measures on the market for or value of copyrighted works."[111] There is little chance[112] that the proposed expansion could "devalue" works or "undermin[e] legitimate markets" as AAP argues.[113] Institutions that build corpora must have or obtain lawful copies of all underlying works in each corpus; institutions that *receive* corpora must *also* have or obtain lawful copies of all the underlying works. Further, uses permitted by the proposed expansion would not interfere with any licensing market in any cognizable manner.[114]

In any case, uses are certainly not likely to be "widespread" as AAP fears.[115] As explained in the Initial Comment, "the expected number of uses is not likely to be substantial,"[116] given the small size and highly specialized nature of the digital humanities field.[117] And as further described above, the robust security measures required in the original exemption still adequately guard against the unauthorized distribution of copyrighted works. This is the exact same requirement that exists in the current exemption, for the exact same non-expressive use, that the Office found not to be a market substitute in the 2021 Triennial Proceeding.[118]

[109] Initial Comment at 18 (quoting App. J: Letter from Rachael Samberg and Timothy Vollmer at 4). *See also* Initial Comment, App. L: Letter from Henry Alexander Wermer-Colan at 1.

[110] 2021 Recommendation at 121.

[111] 17 U.S.C. § 1201(a)(1)(C)(iv).

[112] *See* Initial Comment at 22–23, 30–32 (discussing lack of market harm).

[113] AAP Class 3(b) Opp'n at 16.

[114] 2021 Recommendation at 112 ("The Register concludes that, with the limitation that researchers may not use the copies of the copyrighted works in the corpus for their expressive purposes, the copies would not serve as substitutes for the original works.").

[115] AAP Class 3(b) Opp'n at 16.

[116] Initial Comment at 24.

[117] *Id*.

[118] 2021 Recommendation at 112.

e. Opponents make no material arguments under the fifth statutory factor.

Opponents make only one argument under the fifth § 1201 factor. AAP contends that the exemption is at odds with the purposes of § 1201 because it discourages copyright owners' reliance on access controls.[119] That same argument could be raised in the consideration of any § 1201 temporary exemption and does not meaningfully influence the analysis here.

[119] AAP Class 3(b) Opp'n at 16.

DOCUMENTARY EVIDENCE

Appendix A: Example of Dataset Preparation and Computational Analysis

Appendix B: Uncropped screenshot of Kinolab

Appendix C: Response from Bowdoin College to the MPA

Appendix A
Example of Dataset Preparation and Computational Analysis

APPENDIX A: EXAMPLE OF DATASET PREPARATION AND COMPUTATIONAL ANALYSIS

This Appendix provides an in-depth example of the painstaking, time-consuming, and research-directed process for preparing works for TDM research; how such work contributes to the growth of the field and enables other research; and why domain-specific corpora are needed. It supplements the multiple examples in the Initial Comment.

In *Measuring Information Propagation in Literary Social Networks* ("***Information Propagation***"), the authors Matthew Sims and David Bamman "describe a new pipeline for measuring information propagation" and "analyze the dynamics of information propagation in over 5,000 works of English fiction."[1] Such analysis requires multiple nontrivial steps: (1) *who*—identifying the parties through whom the information flows; (2) *what*—defining what the information is; and (3) *how*—determining how information is propagated.[2] Furthermore, this analysis must be conducted in such a way that computers are able to read and analyze the data. The result of such labor is a dataset of quotations alongside their speakers, and this dataset is then leveraged to discover trends in English language literature. For instance, findings from this corpus include claims such as "women propagate information *between* men much more rarely than they do in other configurations"; this claim builds upon previous literary criticism that "envision[s] the role of women in novels as being intermediaries between men."[3]

(1) *Who*—identifying the parties the information flows through. Characters are not referenced in only a single way within a piece of text. Character references may occur with or without honorifics, as a pronoun, as a nickname, or as a descriptor. For example, in *The Brothers Karamazov,* Dmitri Fyodorvich Karamazov is referred to as Dmitri, Mitya, Mitri, Mr. Karamazov, I, you, he, him, the prisoner, the murderer, your brother, my son, and others.[4] These are all distinct strings of text to a computer, and the dozens of dozens of references must be all correctly mapped to the dozens of characters. Before such mapping can happen**,** however, a computer must know when a piece of text is even referring to a character, rather than a place or a verb. *Information Propagation* was able to rely on work by authors of a previous article, *An Annotated Dataset of Literary Entities*, who built a dataset from 100 novels and annotated information such that a machine learning model is able to detect when an entity appears in text and what kind of entity it is (e.g., people, facilities, or geopolitical entities).[5] With this, another set of authors, in *An Annotated Dataset of Coreference in English Literature*, were able to annotate the same 100-novel

---

[1] Matthew Sims & David Bamman, *Measuring Information Propagation in Literary Social Networks*, *in* Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) 642 (Bonnie Weber et al. ed., 2020), https://aclanthology.org/2020.emnlp-main.47/.

[2] *Id*. at 644.

[3] *Id*. at 649.

[4] Fyodor Dostoevsky, The Brother Karamazov (Constance Garnett trans., 1912) (1880).

[5] David Bamman et al., *An Annotated Dataset of Literary Entities*, *in* 1 Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Long and Short Papers) 2138 (Jill Burstein et al. ed., 2019), https://aclanthology.org/N19-1220/.

dataset,[6] with all three authors annotating coreferences, so that their model could determine when two pieces of text reference the same character.[7] Because the previous work existed, the authors of *An Annotated Dataset of Coreference in English Literature* were not only able to build upon it and provide additional annotations to "investigate the temporal distance over which entities exist in discourse," but could also, through sharing the annotations, provide a solid foundation for future work such as *Information Propagation*.[8]

(2) *What*—defining what the information is. In order to measure how information is being propagated, the authors used dialogue as a representation for the propagated information. This involves algorithmically identifying when conversation is occurring in a text and who is speaking. The first task is no trivial endeavor either. While quotation marks serve as a helpful indicator,[9] the authors had to account for multiple types of errors: "scare quotes," titles, and speech that does not use quotation marks at all, such as in James Joyce's *Ulysses*.[10] Further work was done to algorithmically attribute quotes to the appropriate speaker. The list of potential speakers was made possible by the previous research and the annotations described in it. Building upon the previous dataset of 100 novels, the annotators working on *Information Propagation* were able to annotate and "identify all quotations and attribute each one to a speaker who uttered it" for the entire 100 texts.[11] To ensure consistency, the authors manually double-checked 10% of the annotations for the 100 novels described in the previous papers.[12] Then, using these annotations, the authors were able to measure the accuracy of their speaker attribution methods.[13] After the dialogue was adequately attributed, represented, and verified, researchers were left with a set of tabular data that takes the following form:[14]

---

[6] Note that this dataset does not contain the entirety of each novel, but 2,000 words from each, to ensure that one extremely long work does not statistically dominate a smaller work. *Id.* at 2139.

[7] David Bamman et al., *An Annotated Dataset of Coreference in English Literature*, *in* Proceedings of the Twelfth Language Resources and Evaluation Conference 44 (Nicoletta Calzolari et al. ed., 2020), https://aclanthology.org/N19-1220/.

[8] *Id*. at 44.

[9] Note that care still needs to be taken at this step, since "straight quotes," 'single directional quotes,' and "double directional quotes" represent six different possible characters. Furthermore, different languages use different methods, requiring additional effort for TDM research conducted on works that are in languages other than English. For example, Chinese, Japanese, and Korean use 「corner brackets」, French uses « guillemets », and German uses quotation marks, but in an „inverse order."

[10] Sims & Bamman, *supra* note 1, at 645.

[11] *Id*. at 644.

[12] *Id*.

[13] *Id*. at 645.

[14] David Bamman, *LitBank*, https://github.com/dbamman/litbank/tree/master/quotations (last visited Mar. 19, 2024). It goes without saying that such a corpus cannot substitute for reading the novel itself.

Label	Quote ID	Start sentence ID	Start token ID (within sentence)	End sentence ID	End token ID (within sentence)	Quotation
QUOTE	Q342	54	0	54	13	“ Of course , we ’ll take over your furniture , mother , ”

Label	Quote ID	Speaker ID
ATTRIB	Q342	Winnie_Verloc-3

Once the quotations and speakers had been identified, it was possible to computationally extract information from the quotation. The authors computationally extracted “propositional tuples of the form (subject, verb, object)” from the text so that, for example, the sentence “Bob punched Tom and he left” resulted in two tuples (Bob-id1, punch, Tom-id2) and (Bob-id1, leave, null).[15] In addition, after engaging in some cleanup (e.g., removing any tuples with first or second pronouns), the authors picked the 100 words that occurred most frequently across all tuples, and placed them into four categories: amorous (e.g., love, marriage), hostile (e.g., hurt, hit, shoot, kill), juridical (e.g., arrest, escape, innocent, guilty), and vital (e.g., alive, sick, dead).[16] Importantly, the frequency of these terms and the subsequent categorizations by the authors is a result of the types of novels the dataset contains, all of which were drawn from Project Gutenberg, which primarily hosts nineteenth-century novels.[17] A corpus using novels spanning a different time period would employ different words in different frequencies; as such, using a model from such a corpus would result in reduced accuracy in this context. Further, out-of-domain techniques such as measuring information repetition (which is sufficient in news contexts due to the substantial repetition of text in those contexts) would be wholly insufficient to the task of analyzing novels.[18]

(3) *How*—defining how information is propagated. After all of this preparation, the authors were finally able to measure how information is spread. The authors categorized two different types of propagation: implicit and explicit. Implicit propagation occurs when a tuple of information gets told from character A to character B and then from character B to character C. Explicit propagation occurs when the characters engage in hearsay, e.g., some variant of “[character-id] said.”[19]

[15] Sims & Bamman, *supra* note 1, at 646.

[16] *Id*.

[17] *Id*.

[18] *See id*. at 645–46.

[19] *Id*. at 646.

From these one hundred works, the authors created a computational model to analyze 5,345 works within the Project Gutenberg corpus.[20] With this analysis, the authors were able to empirically support a version of Granovetter's theory of weak ties, namely that "information in novels propagates through characters that serve as bridges between otherwise disconnected communities."[21] In addition, the authors were able to support the thesis that women are stereotyped to engage in more gossip. Specifically, they showed that "not only are female characters more likely to serve as propagators than male characters in this dataset, but that female characters fill this role more frequently than one would expect given the proportion of [female intermediaries in all A-B-C character triplets]."[22]

* * *

At every step, additional annotations had to be created for the 100 different literary novels in a dataset of 210,532 tokens.[23] These annotations are closely tied to the underlying text.[24] Creating the original dataset in *An Annotated Dataset of Literary Entities* involved annotating entities so that objects such as people, facilities, geopolitical entities and more are labeled.[25] In *An Annotated Dataset of Coreference in English Literature*, the authors annotated coreferences so that a model could learn when pieces of texts are referring to the same character.[26] The authors of *Information Propagation* then annotated quotations and their speakers.[27] Each built on the previous. Notably, such a model is necessary because while other machine learning models do exist, the domains of those systems (e.g., news, conversation, the Bible, and the Web) are insufficient to capture the nuances and idiosyncrasies of literary texts.[28] Because of this previously existing annotated corpus, the authors of *Information Propagation* were able to ask new questions and build upon previous work.

Note: this example is still *simpler* than the TDM research projects at issue in this proceeding, because the researchers studied public domain materials already available as plaintext files; as such, the researchers neither had to undertake the technical process of circumventing TPMs nor had to worry about incurring liability for engaging in that circumvention. They were "simply" able to engage in noninfringing, socially beneficial research, without confronting those additional barriers.

---

[20] *Id*. at 647.

[21] *Id*. at 648.

[22] *Id*. at 649.

[23] Tokens roughly correspond to words. *See* Dan Jurafsky & James H. Martin, *Speech and Language Processing* 19 (Feb. 3, 2024) (unpublished manuscript), https://web.stanford.edu/~jurafsky/slp3/ed3bookfeb3_2024.pdf.

[24] *See* Bamman et al., *An Annotated Dataset of Literary Entities*, *supra* note 5, at 2138; Bamman et al., *An Annotated Dataset of Coreference in English Literature*, *supra* note 7, at 44.

[25] Bamman et al., *An Annotated Dataset of Literary Entities*, *supra* note 5, at 2139.

[26] Bamman et al., *An Annotated Dataset of Coreference in English Literature*, *supra* note 7, at 44.

[27] Sims & Bamman, *supra* note 1, at 642.

[28] Bamman et al., *An Annotated Dataset of Coreference in English Literature*, *supra* note 7, at 44.

Appendix B
Uncropped screenshot of Kinolab

Register Sign in

Kinolab

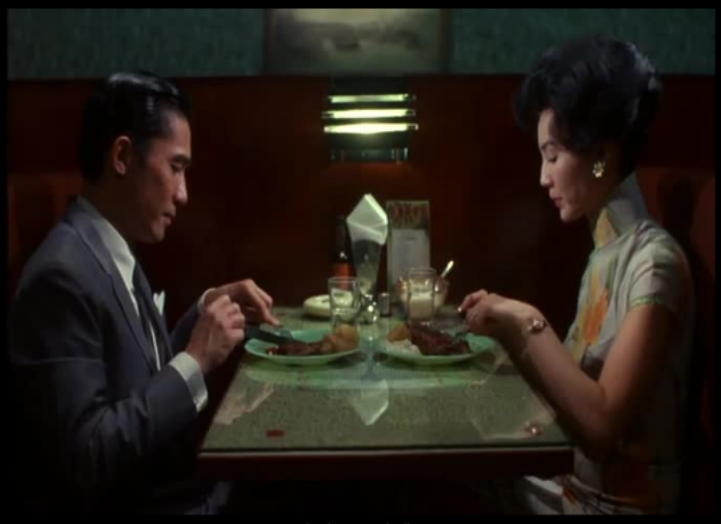
Collection Research Teaching Collaborations Sign in About

Search clips
Custom search

Films

In the Mood for Love : *A Dinner With Mr. Chan and Mrs. Chow*

Mr. Chow (Tony Leung) and Mrs. Chan (Maggie Cheung) pretend to have dinner together as the other's spouse.



Sign in to watch clips

Tags Citation Contributor(s)

Film Language

180-degree rule, close-up, costume, dolly shot, establishing shot, framing, high angle, long shot, long take, medium close-up, music, diegetic, music, nondiegetic, over the shoulder shot, pan, shallow focus, shot/reverse shot, two-shot

Diversity in Media Representation

Bowdoin Cinema Studies

Bowdoin Digital and Computational Studies

Appendix C
Response from Bowdoin College to the MPA

Bowdoin

February 15, 2024

Karyn A. Temple
Senior Executive Vice President, Global General Counsel
Motion Picture Association
Karyn_Temple@motionpictures.org

Dear Ms. Temple:

I am writing in response to your letter of January 31, 2024, to Professor Allison Cooper, director of Kinolab, a digital humanities laboratory for the analysis of film language at Bowdoin College. Professor Cooper outlined Kinolab's research activities and use of films and television within the confines of the current TDM exemption in her letter in support of the Long Comment Regarding a Proposed Exemption Under 17 U.S.C. § 1201 submitted by the Authors Alliance, the American Association of University Professors, and the Library Copyright Alliance, which is attached here for your reference.

We have reviewed the MPA's request for information under 37 C.F.R. § 201.40(b)(4)(ii)(B). (You reference 37 C.F.R. § 201.40(b)(5) in your letter, which pertains to literary works, not motion pictures. We assume that you did so in error and intended to reference 37 C.F.R. § 201.40(b)(4) instead.) The term "highly confidential information" is undefined in 37 C.F.R. § 201.40(b)(4)(ii)(B) and is in general an ambiguous term without a definitive meaning, making it difficult to respond to your information request, particularly in light of the short response timeframe you have provided.

That said, we attach Bowdoin College's Written Information Security Program and Data Classification Policy, both of which are approved by Bowdoin College's Chief Information Officer. These policies provide the MPA with a sound overview of the measures that Bowdoin College uses to keep its own "highly confidential information" secure. Additional information regarding Bowdoin College's data security policies and standards can be found on its publicly available website at https://www.bowdoin.edu/it/safe-computing/data-security-policies.html. These policies are communicated to all members of the College community, including to those researchers and other institutional actors engaged in circumvention activities under the Exemption.

We expect that this response satisfies the MPA's inquiry. However, you may direct any further inquiries concerning this matter to my attention.

Sincerely,

Megan A. Hart

Megan A. Hart, Esq.
Legal Officer

Office of Legal Counsel
5600 College Station, Brunswick, ME 04011-8447 207.725.3092 mhart@bowdoin.edu

Written Information Security Program (WISP)

Policy

Authority

This policy is approved by the Chief Information Officer.

Purpose

This policy was implemented to identify and protect against potential information security risks at the institution and to comply with regulations issued by various states. This program has been developed in accordance with the following security best practices and regulations:

- National Institute for Standards and Technology (NIST) Cybersecurity Framework (CSF)
- State of Maine Title 10, §1348. Security breach notice requirements
- Payment Card Industry Data Security Standards (PCI-DSS)
- Graham-Leach-Bliley Act (GLBA or GLB Act)
- Family Educational Rights and Privacy Act (FERPA)
- Health Insurance Portability and Accountability Act (HIPAA)
- 201 CMR 17: Standards for the Protection of Personal Information of MA Residents
- Massachusetts General Laws Chapter 93H

The purpose of this policy is to:

- Establish a comprehensive information security program for Bowdoin College with policies designed to safeguard sensitive data that is maintained by the College, in compliance with federal and state laws and regulations.
- Establish administrative, technical and physical safeguards to ensure the security of sensitive data that aligns with current best practices.
- Ensure clear communication of information security policies and standards.

Bowdoin is committed to protecting the confidentiality of all sensitive data that it maintains. Bowdoin has implemented policies to protect such information, and the WISP should be read in conjunction with these policies that are cross-referenced at the end of this document.

1. Scope

This Program applies to all Bowdoin employees, whether full- or part-time, including faculty, administrative staff, union staff, contract and temporary workers, hired consultants, interns, and student employees, as well as to all other members of the Bowdoin community (hereafter referred to as the "Community"). This program also applies to contracted third-party vendors. The systems and data covered by this Program includes any information stored, accessed, or collected at the College or for College operations. The WISP is not intended to supersede any existing Bowdoin College policy that contains more specific requirements for safeguarding certain types of data, except in the case of Personally Identifiable Information as defined below. If such policy exists and it conflicts with the requirements of the WISP, the other policy takes precedence.

2. Definitions

NIST CSF - The National Institute for Standards and Technology publishes the Cybersecurity Framework as a guide to help organizations with the security of critical infrastructure. See: https://www.nist.gov/cyberframework

Data - For the purposes of this Program, data refers to any information collected, accessed and stored about members of the College community.

Personally Identifiable Information ("PII") - is the first name and last name or first initial and last name of a person in combination with any one or more of the following:

- Social Security number.
- Driver's license number or state-issued identification card number.
- Financial account number (e.g. bank account) or credit or debit card number that would permit access to a person's financial account, with or without any required security code, access code, personal identification number, or password.

For the purposes of this Program, PII also includes passport number, alien registration number or other government-issued identification number.

Nonpublic Financial Information - The GLBA (FTC 16 CFR Part 313) requires the protection of "customer information", that applies to any record containing nonpublic financial information ("NFI") about a student or other third party who has a relationship with the College, whether in paper, electronic or other form, that is handled or maintained by or on behalf of the College. For these purposes, NFI shall include any information:

- A student or other third party provides in order to obtain a financial product or service from the College.
- About a student or other third party resulting from any transaction with the College involving a financial product or service.
- Otherwise obtained about a student or other third party in connection with providing a financial product or service to that person.

Examples of NFI include:

- Information a consumer provides to you on an application to obtain a loan, credit card, or other financial product or service.
- Account balance information, payment history, overdraft history, and credit or debit card purchase information.
- The fact that an individual is or has been one of your customers or has obtained a financial product or service from you.
- Any information about your consumer if it is disclosed in a manner that indicates that the individual is or has been your consumer.
- Any information that a consumer provides to you or that you or your agent otherwise obtain in connection with collecting on, or servicing, a credit account.
- Any information you collect through an Internet "cookie" (an information collecting device from a web server).
- Information from a consumer report.

Breach - the unauthorized acquisition or unauthorized use of unencrypted data or, encrypted electronic data and the confidential process or key that is capable of compromising the security, confidentiality, or integrity of personal information, maintained by a person or agency that creates a substantial risk of identity theft or fraud against a resident of the commonwealth. A good faith but unauthorized acquisition of personal information by a person or agency, or employee or agent thereof, for the lawful purposes of such person or agency, is not a breach of security unless the personal information is used in an unauthorized manner or subject to further unauthorized disclosure.

3.0 Roles and Responsibilities

Senior Officers - Responsible for making a final review and approval of this policy.

IT Governance - Responsible for reviewing and providing feedback prior to final approval by Senior Officers.

Chief Information Officer - Responsible for reviewing and approving this policy prior to Executive Leadership. The CIO must report all compliance-related activities pertaining to this policy to the Executive Leadership team.

Chief Information Security Officer - Responsible for creating policies, procedures and standards to meet these established requirements as outlined in this policy. The ISO must report all matters pertaining to compliance with this policy to IT Governance.

Data Governance - Partnership in reviewing, developing and disseminating critical information security policies and standards.

4.0 Information Security Operations

4.1 Data Classification Policy

All data covered by this policy is subject to the Colleges Data Classification Policy. The purpose of this policy is to protect the information resources of the College from unauthorized access or damage. The requirement to safeguard information resources must be balanced with the need to support the pursuit of legitimate academic objectives. The value of data as an institutional resource increases through its widespread and appropriate use; its value diminishes through misuse, misinterpretation, or unnecessary restrictions to its access. By default, all institutional data will be designated as "Sensitive".

All data at the College is assigned a data trustee according to the constituency it represents. Data Trustees are senior college officials or their designees who have planning, policy-level and management responsibility for data within their functional areas. Data Stewards are college officials having direct operational-level responsibility for the management of one or more types of data. Data Stewards are assigned by the Data Trustee and are generally associate deans, associate vice presidents, directors, or managers. One of the responsibilities of the Data Governance Committee is to maintain a data inventory of the College's data assets, which lists the Data Steward(s) for each data set. Please see the Data Governance Committee website to find the current information on the various roles.

4.1.1 Personally Identifiable Information

Additional security controls and safeguards will be implemented to ensure the security and confidentiality of PII. Safeguarded PII includes community, student and employee information. Bowdoin will assess the security controls in place to safeguard PII data on an annual basis to ensure the following:

- The implemented security controls are appropriate based on the type of the business and the need to safeguard the PII.
- The implemented security controls are operating effectively to prevent the unauthorized use and access of personal information
- Adequate resources and staff are available to ensure the security of PII
- The appropriate controls are in place based on the amount of stored data
- Information safeguards are upgraded as necessary to limit security risks.

Please refer to section 4.7 for further details on security monitoring and assessment.

4.2 Information Security Standards and Policies

The Information Security Policies are guiding documents to ensure that Bowdoin keeps the information that it creates or processes secure during regular operations.

The Information Security Standards are to be used when implementing operational security at the College. These standards should be used when deploying technology for the college or protecting information.

All security policies and standards shall be reviewed and approved annually by the CIO and CISO. Security policies and standards will also be reviewed to address any sizable business change, and/or to incorporate lessons from incident response.

4.3 Acceptable Use Policy

The College has several policies relating to the acceptable use of technology. Users of Bowdoin College network and computer resources have a responsibility to properly use and protect those information resources and to respect the rights of others. Please see the following policies:

- Computer and Network Usage Policy
- Business Computing Policy
- Information Technology Computer Use Policy

All Bowdoin information systems will be monitored for unauthorized uses or access to personal information. Please refer to section 4.7 for additional details on security monitoring and assessment.

4.4 Encryption Standards

- Please see the IT Security Standard: Encryption for the complete encryption standard
- Encryption of data at rest- all Bowdoin employee laptops and desktops are encrypted. Any removable storage media such as external hard drives or usb storage devices must be encrypted if it stores sensitive data.
- Encryption of Data in transit- the transmission of all sensitive data will be encrypted in transit using secure encryption algorithms and methodologies. Transmitted data includes data traveling across public networks and wirelessly. Bowdoin employees that are accessing the College's internal network resources remotely will only use encrypted VPN connections per the Virtual Private Network (VPN) Policy.

4.5 Incident Response

If there is a breach that requires notification under state or federal law, Bowdoin shall follow the appropriate incident response process outlined in our Incident Response Plan. The Bowdoin Incident Response plan is maintained by the Chief Information Security Officer and includes coordination with the other IT groups.

Incident Reporting:

If a breach or incident is suspected, please contact the security team for further support and guidance at IT Security at itsecurity@bowdoin.edu or by phone at 207-798-4248. If the security team is unreachable, please contact the Service Desk at servicedesk@bowdoin.edu or 207-725-3030.

Post Incident Review: The Bowdoin Incident Response Team will document any responsive actions taken if a security incident involves a data breach. The review will include assessment of the effectiveness of the current security controls and will identify potential technical and business changes to prevent and mitigate security incidents in the future.

4.6 Physical Security

Bowdoin shall implement physical security controls to protect sensitive information such as PII and NFI. Physical security safeguards include but are not limited to the secure storage of records, facilities or containers containing PII.

4.7 Security Monitoring

Access to Bowdoin systems is monitored regularly and the process is outlined in the IT Security Standard: Log Management and Security Monitoring.

The Bowdoin information security program will be monitored to ensure the appropriate security measures and controls are implemented and operating effectively to ensure the protection of sensitive data from unauthorized access and use. Internal assessment is done annually to assess controls against the NIST CSF. Annually external partners are brought in to further validate and assess the appropriateness of controls that have been implemented.

4.8 Risk Management

To successfully manage risk across Bowdoin College, senior leadership must be committed to making information security an underlying principle of operating the College to protect the institution and its community. This top-level commitment ensures that sufficient resources are available to develop and implement an effective, institution-wide security program. Effectively managing information security risk requires the following key elements:

- Assignment of risk management responsibilities to appropriate senior leadership.
- Ongoing recognition and understanding by senior leadership and IT Governance of the information security risks to Bowdoin information assets, operations, and personnel.
- Establishment of the tolerance for risk and communicating the risk tolerance throughout the organization, including guidance on how risk tolerance impacts ongoing decision-making activities.
- Providing accountability for senior leadership for their risk management decisions.
- Ongoing assessments of internal and external risks

4.8.1 Risk Assessments

Bowdoin recognizes that it has both internal and external risks to the security, integrity and confidentiality of College information. These risks include, but are not limited to:

- Unauthorized access of confidential data
- Compromised system security as a result of unauthorized access
- Interception of data during transmission
- Loss of data integrity
- Physical loss of data in a disaster
- Errors introduced into the system
- Corruption of data or systems
- Unauthorized access of confidential data by employees
- Unauthorized requests for confidential data
- Unauthorized access through hard copy files or reports
- Unauthorized transfer of confidential data through third parties
- Employee compliance with security training and security policies and standards

This is not a complete list of risks and the threat landscape will evolve during the life of the WISP. This list of risks will be updated and reassessed at least annually. The Information Technology team must implement reasonable and sufficient safeguards to provide security and confidentiality to confidential data maintained by the College.

4.8.2 Third Party Risk

All third-party service providers shall be subject to a security risk review prior to entering into an agreement and on a regular basis. The security team will review the security controls that the Third Party has in place to ensure are consistent with applicable state and federal regulations and Bowdoin Colleges security policies. Compliance of security controls will be mandated through contractual requirements.

This review is performed using the Higher Education Community Vendor Assessment Toolkit (HECVAT) developed by EDUCAUSE. Any vendors that handle Restricted data under our data classification policy will be required to fill out the full assessment, vendors that handle less sensitive data may use the shorter versions of the assessment.

4.9 Employee Training

All administrative employees are required to complete annual information security awareness training. This is achieved through completion of a formal awareness training course and may also be accompanied by additional awareness materials as a part of National Cybersecurity Awareness Month. This course must be completed annually, or access to information resources may be revoked. In the event of a successful phishing attack, remedial training may be required regardless of whether the annual training has already been completed.

4.11 Identification, Authentication and Authorization

User authentication and authorization protocols and passwords are secured using encryption. All information system identifiers are unique, and one identifier is used per individual. Only approved, secure authentication mechanisms are allowed. Management of system accounts and authentication mechanisms follow the Identification, Authentication, and Authorization Policy.

All system account passwords must use a strong, encrypted password in accordance with the Bowdoin Passwords Security Standard.

4.11.1 Remote Access

All employees must abide by Bowdoin's security policies and standards when accessing Bowdoin's information systems remotely. Encrypted connections must be used to access information systems containing PII or other types of sensitive data. Employees are forbidden to use personal devices to access or store PII or Restricted data from Bowdoin's information systems.

4.12 System Maintenance

All information systems, operating systems and software will be patched on a regular basis in accordance with the IT Security Standard: System Maintenance. In addition to patching, virus definitions will be updated on a regular basis for all anti-malware solutions.

5.0 Enforcement

A user of College information resources who is found to have purposely or recklessly violated any of these policies, or who fails to comply with this Program in any other respect, will be subject to disciplinary action according to the policies of Human Resources or the Dean of Student Affairs office, up to and including discharge, dismissal, expulsion and / or legal action.

6.0 Related Policies

Bowdoin has adopted these policies and standards to support this program:

- Data Classification Policy
- Computer and Network Usage Policy
- Business Computing Policy
- Virtual Private Network (VPN) Policy
- Email Policy
- Information Technology Computer Use Policy
- Identification, Authentication, and Authorization Policy

- IT Security Standard: Bowdoin Passwords
- IT Security Standard: Managed Computer/Endpoint Configuration
- IT Security Standard: Logging and Security Monitoring
- IT Security Standard: Encryption
- IT Security Standard: System Maintenance and Vulnerability Management

7.0 Implementation

Effective Date	March 15, 2023
Review Frequency	Annual
Responsible Officer	Chief Information Officer

Sign in to leave feedback

0 reviews

Details

Article ID: 154310
Created
Mon 3/13/23 10:23 AM

Modified
Wed 3/15/23 11:50 AM

Attachments (0)

No attachments found.

Data Classification Policy

Policy

Authority

This policy is approved by the Chief Information Officer (CIO).

Summary

All College data is classified into defined access levels. Data may not be accessed without proper authorization.

The purpose of this policy is to protect the information resources of the College from unauthorized access or damage. The requirement to safeguard information resources must be balanced with the need to support the pursuit of legitimate academic objectives. The value of data as an institutional resource increases through its widespread and appropriate use; its value diminishes through misuse, misinterpretation, or unnecessary restrictions to its access.

1. Classification of Data

All College data is classified into levels of sensitivity to provide a basis for understanding and managing college data. Accurate classification provides the basis to apply an appropriate level of security to college data. These classifications of data take into account the legal protections (by statute, regulation, or by the data subject's choice), contractual agreements, ethical considerations, or strategic or proprietary worth. Data can also be classified as a result of the application of "prudent stewardship", where there is no reason to protect the data other than to reduce the possibility of harm or embarrassment to individuals or to the institution.

By default, all institutional data will be designated as "Sensitive". College employees will have access to the data for use in the conduct of college business.

2. Classification Levels

The classification level assigned to data will guide Data Stewards, Data Managers, business and technical project teams, and any others who may obtain or store data, in the security protections and access authorization mechanisms appropriate for that data. Such categorization encourages the discussion and subsequent full understanding of the nature of the data being displayed or manipulated. Data is classified as one of the following:

- **Public (low level of sensitivity)**
  Access to "Public" institutional data may be granted to any requester. Public data is not considered confidential. Examples of Public data include published directory information and academic course descriptions. The integrity of Public data must be protected, and the appropriate Data Manager must authorize replication of the data. Even when data is considered Public, it cannot be released (copied or replicated) without appropriate approvals.
- **Sensitive (moderate level of sensitivity)**
  Access to "Sensitive" data must be requested from, and authorized by, the Data Steward who is responsible for the data. Data may be accessed by persons as part of their job responsibilities. The integrity of this data is of primary importance, and the confidentiality of this data must be protected. Examples of Sensitive data include purchasing data, financial transactions that do not include restricted data, information covered by non-disclosure agreements and Library transactions.
- **Restricted (highest level of sensitivity)**
  Access to "Restricted" data must be controlled from creation to destruction, and will be granted only to those persons affiliated with the College who require such access in order to perform their job, or to those individuals permitted by law. The confidentiality of data is of primary importance, although the integrity of the data must also be ensured. Access to restricted data must be requested from, and authorized by, the Data Steward who is responsible for the data. Restricted data includes information protected by law or regulation whose improper use or disclosure could:

  - Adversely affect the ability of the college to accomplish its mission
  - Lead to the possibility of identity theft by release of personally identifiable information of college constituents
  - Put the college into a state of non-compliance with various state and federal regulations such as FERPA, HIPAA, and GLBA
  - Put the college into a state of non-compliance with contractual obligations such as PCI DSS

  The specification of data as restricted should include reference to the legal or externally imposed constraint that requires the restriction, the categories of users typically given access to the data, and under what conditions or restrictions access is typically given.

  Examples of Restricted data include social security numbers, student registration, grades, financial aid data and bank account numbers.

3. Roles and Responsibilities

Chief Information Security Officer

The Chief Information Security Officer implements policies and procedures to comply with the Health Insurance Portability and Accountability Act of 1996 (HIPPA), Family Education Rights and Privacy Act (FERPA), and others governing the treatment of individually identifiable information.

Data Trustees

Data Trustees are senior college officials or their designees who have planning, policy-level and management responsibility for data within their functional areas. Data Trustees responsibilities include:

- Assigning and overseeing Data Stewards
- Remaining aware of the legal and regulatory requirements for data in their areas
- Ensuring that data policies are established, and kept up to date, in their areas and if appropriate, delegating such responsibility
- Promoting appropriate use and data quality

Data Stewards

Data Stewards are college officials having direct operational-level responsibility for the management of one or more types of data. Data Stewards are assigned by the Data Trustee and are generally associate deans, associate vice presidents, directors or managers. Data Steward responsibilities include:

- The application of this and related policies to the systems, data, and other information resources under their care or control
- Overseeing the establishment of data policies in their areas
- Understanding legal and regulatory requirements for data in their areas
- Classifying data using the College's data classification system
- Identifying safeguards for Restricted Data

In cases where multiple Data Stewards collect and maintain the same restricted data elements, the Data Stewards must work together to implement a common set of safeguards.

Data Managers

Data Managers are college officials who are responsible for day-to-day operational data collection and management, overseeing the life cycle of a particular set of institutional data. They have the authority from the Data Steward and/or Data Trustee to grant internal access to data for their functional area. Data Managers are generally managers of data systems or senior data analysts within business departments. Data Manager responsibilities include:

- Implementing the established data policies in their areas
- Developing data definitions and standards for data elements in their functional area
- Regularly striving to improve the way data is defined, produced, and used in their functional area
- Resolving data quality issues pertaining to data in their functional area
- Safeguarding data by ensuring appropriate access, following established authorization procedures, and maintaining physical and system security appropriate to the classification level of the data in their custody
- Following data handling and protection policies and procedures established by Data Stewards and Information Security
- Communicating and providing education on the required minimum safeguards for protected data to authorized data users
- Supporting access by providing appropriate documentation and training to data consumers

- Setting an example of data-related behavior for their department

Data Consumers

Data Consumers are the individual college community members who have been granted access to college data in order to perform assigned duties or in fulfillment of assigned roles or functions at the college. This access is granted solely for the conduct of college business. Data Consumer responsibilities include:

- Following the policies and procedures established by the relevant Data Steward and Information Security
- Complying with federal and state laws, regulations, and policies associated with the college data used
- Applying safeguards prescribed by appropriate Data Steward for Restricted Data

Reporting any unauthorized access or data misuse to Information Security or the appropriate Data Steward for remediation.

Details

Article ID: 71630

Created

Thu 2/7/19 10:31 AM

Modified

Fri 2/10/23 11:45 AM

Attachments (0)

No attachments found.