



## A2P2 ISSUE BRIEF: PETITION FOR EXEMPTION TO SECTION 1201 TO ENABLE TEXT AND DATA MINING RESEARCH

MARCH 2021

### ABOUT SECTION 1201 AND THE PROPOSED EXEMPTION FOR TDM RESEARCH

Authors Alliance, joined by the Library Copyright Alliance and the American Association of University Professors, are petitioning for a new exemption to section 1201 of the Digital Millennium Copyright Act (“DMCA”) that would allow researchers to bypass technical protection measures (“TPMs”) on literary works distributed electronically and motion pictures for the purpose of conducting text and data mining (“TDM”) research.

Section 1201 of the DMCA prohibits the circumvention of TPMs used by copyright owners to control access to their works. It also prohibits the manufacture or sale of devices or programs designed to circumvent these TPMs. In other words, section 1201 prevents individuals from breaking digital locks on copyrighted works, even when they seek to make a fair use of those copyrighted works or engage in otherwise non-infringing activities. Because section 1201’s prohibitions can interfere with fair and socially beneficial uses of copyrighted works, the DMCA also provides for a triennial rulemaking process to grant temporary exemptions to the prohibitions.

**Read More:** [Comment in Support of Exemption for TDM Research](#), Authors Alliance  
[Rulemaking Proceedings Under Section 1201 of Title 17](#), Copyright Office

### HOW DOES SECTION 1201 IMPEDE TDM RESEARCH?

Text and data mining refers to automated techniques aimed at analyzing digital text and data in order to generate information that reveals patterns, trends, and correlations in that text or data. TDM has great potential to enable groundbreaking research and contribute to the commons of knowledge. As a highly transformative use of copyrighted works done for purposes of research and scholarship, TDM fits firmly within the ambit of fair use. But the current prohibition on bypassing TPMs in section 1201 makes TDM research on texts and films time consuming and inefficient—and in some cases, impossible—working against the promotion of the progress of knowledge and the useful arts that copyright law has been designed to incentivize.

Because literary works distributed electronically and motion pictures are protected TPMs, researchers—unable to bypass these TPMs due to section 1201—typically engage in their research by

manually scanning physical books and using optical character resolution on the resulting digital files to create minable versions and by using screen capture technology on films. However, these techniques are time-consuming, often error-ridden, and impossible to deploy at scale. TDM researchers may also turn instead to works in the public domain for their TDM research. With regard to films, this avenue is effectively unavailable, since works published after 1925 generally remain under copyright. For literary TDM scholars, literary works published before 1925 remain a potential alternative area of study, but focusing TDM on pre-1925 texts further reinscribes white men as the center of the field and further marginalizes women and people of color. Authorship was far less diverse in 1925 than it is today, so TDM research on public domain texts ends up privileging white male voices rather than being representative of authors contributing to the commons of knowledge today.

**Read More:** [Law and Literacy in Non-Consumptive Text Mining: Guiding Researchers Through the Landscape of Computational Text Analysis](#), Rachael G. Samberg and Cody Hennesy  
[Letters in Support of Proposed Exemption for TDM Research](#)

## OPPOSITION TO THE PETITION

In February, four comments were submitted in opposition to the proposed new exemption. Opposition commenters raised concerns about the scope of activities and works that would be covered by the exemption as well as the intended beneficiaries of the exemption. One animating concern was that a broadly constructed exemption would permit use of literary works for artificial intelligence and machine learning purposes, including by commercial enterprises, and that such uses could result in the generation of new works that function as market substitutes for the original works that were copied. Opposition commenters also highlighted perceived security risks associated with the creation of full-text databases of decrypted copies of copyrighted works, arguing that the collected works would be exposed to unauthorized downloading and distribution over the internet.

**Read More:** Opposition Comments from [the Motion Picture Association, the Alliance for Recorded Music, and the Entertainment Software Association](#)  
Opposition Comments from [the Association of American Publishers](#)  
Opposition Comments from [the DVD Copy Control Association and the Advanced Access Content System Licensing Administrator](#)

## ADDRESSING OPPONENTS' CONCERNS

Authors Alliance and co-petitioners addressed opposition commenter's concerns in a reply comment submitted in March. Petitioners explained the exemption is intended to cover using text and data mining techniques for the purposes of scholarly research and teaching only and that the intended beneficiaries of the exemption are researchers affiliated with a nonprofit library, archive, museum, or institution of higher education, not commercial actors. With regard to the scope of the works covered, petitioners specified that for the purpose of the exemption, "literary works" would exclude computer

programs. Petitioners also clarified that the exemption will require researchers to take reasonable security measures to ensure that there is no unauthorized access.

**Read More:** [Update: Response Comment in Support of New 1201 Exemption to Enable Text and Data Mining Research](#), Authors Alliance

## WHAT'S NEXT?

In April, stakeholders, including Authors Alliance, will provide testimony at public hearings to help the Copyright Office evaluate the merits of a new exemption to facilitate TDM research. The Librarian of Congress is expected to issue a final decision on the proposed exemption in October 2021.