Academic Researchers Seek to Unlock Copyrighted Books and Films for Text and Data Mining in First-Ever Request to U.S. Copyright Office

Groups say access to contemporary works will aid scholarship about race and gender equity

CONTACTS:

Catherine Crump, Director, Samuelson Law, Technology & Public Policy Clinic, UC Berkeley Law School; Ph. (510) 292-6860; ccrump@law.berkeley.edu

Brianna Schofield, Executive Director, Authors Alliance; Ph. (510) 499-7913; brianna@authorsalliance.org

(BERKELEY, CA, Dec. 15) More than a dozen academic researchers this week filed letters of support for the first-ever petition with the U.S. Copyright Office seeking exemption from liability for breaking digital locks so they can apply sophisticated text and data mining techniques to copyrighted movies, TV shows and literary works.

The Samuelson Law, Technology & Public Policy Clinic at UC Berkeley Law School represents the Authors Alliance, which, along with the Library Copyright Alliance and the American Association of University Professors, is requesting an exemption to Section 1201 of the Digital Millennium Copyright Act (DMCA), the anti-circumvention provision that prevents researchers from working with copyrighted materials that are under a digital lock and key. Violators of the Act face stiff criminal and civil penalties of up to $500,000 and up to 5 years in jail for the first offense and double the fines and jail time for the second offense.

"Because of the DMCA prohibition, data mining techniques cannot be used by scholars to focus on issues that are of even remote contemporary relevance, from the rise of the Harlem Renaissance to the shifting portrayals of Muslim Americans after the 9/11 terrorist attacks to the politics of representation in Network era television," said Brianna Schofield, a copyright attorney and executive director of the Berkeley-based Authors Alliance, whose mission is to empower authors in the digital age. "The prohibition also distorts whose work is analyzed using text data mining by limiting it to the demographic groups most published nearly 100 years ago, which is largely white and male," she added.

If the petition is successful, an exemption will open up a whole new field of research into the motion pictures and literary works created during most of the 20th century and all of the 21st. The Copyright Office considers such petitions every three years and has previously allowed limited exemptions to allow certain categories of users to break digital locks on literary works and motion pictures to make them more accessible to people with disabilities.

"Academic researchers believe text and data mining will allow them to draw new insights from larger collections of texts and films than a person could study in a lifetime," said Catherine Crump, Director of the Samuelson Clinic. "This kind of research poses no threat to the legitimate interests of authors or filmmakers and is the sort of use the Copyright Office's exemption process is designed to protect."

Fourteen academic researchers as well as two experts in academic publishing and the Association for Computers and the Humanities, a professional organization, submitted letters supporting the

petition. The letters from the researchers describe in detail a wide range of projects they hope to pursue using copyrighted material, such as:

- An investigation into whether and how directorial "style" can be measured in films;
- A course in text analytics of African American writers, many of whose works are often underrepresented in available archives;
- A study of novels in which characters are explicitly identified as persons with autism or Asperger syndrome, which is limited to works published after 1994 when these diagnoses were first formally recognized in the DSM-IV.

The petition is online [here](#).

The letters of support are online [here](#).