

**BEFORE THE UNITED STATES COPYRIGHT OFFICE
LIBRARY OF CONGRESS**

**COMMENT REGARDING SECTION 512 STUDY
DOCKET NO. 2015-7**

**COMMENT OF ADAM HOLLAND AND CHRISTOPHER T. BAVITZ
ADDRESSING QUESTIONS 11, 15, 25, 29, AND 30 (AND ALL QUESTIONS GENERALLY)**

I. INTRODUCTION.

Adam Holland and Christopher T. Bavitz (collectively, “Commenters”) submit this comment pursuant to the United States Copyright Office’s Notice of Inquiry for Section 512 Study.¹ Mr. Holland is a Project Coordinator at the Berkman Center for Internet & Society at Harvard University (the “Berkman Center”) and oversees the day-to-day operations of the Center’s Lumen Project (“Lumen”). Professor Bavitz is a faculty co-director of the Berkman Center and the principal investigator for Lumen.

Lumen is an independent research project that studies the landscape for online content, including requests to platforms, search engines, and others to remove materials created or uploaded by Internet users using legal or extra-legal theories. Formed in or around 2001 as the Chilling Effects Clearinghouse, the project’s goals are to educate the public; to facilitate research about different kinds of complaints and requests for removal — both legitimate and questionable — that are sent to online publishers and service providers; and to provide as much transparency as possible about the “ecology” of such notices in terms of who sends them, why, and to what effect.

Commenters write in order to advance the twin propositions that:

- (a) data is crucial to informing reasoned policy debates, including debates about policies that govern intermediary liability and obligations to police content online; and
- (b) transparency is intrinsically related to accountability, oversight, and process and is generally good for the public at large in a society that values free expression.

Such calls for data and transparency around notice-and-takedown stand in contrast to: (a) calls from some quarters that notice data not be publicly shared;² and (b) recent efforts by some platforms to create regimes similar to the DMCA’s notice-and-takedown regime through private ordering, absent transparency for users about the contours of the process.³

¹ 80 Fed. Reg. 81,862 (Dec. 31, 2015).

² Statement of Copyright Alliance CEO Sandra Aistars to H. Judiciary Subcomm. on Courts, IP & Internet Regarding 17 U.S.C. § 512, COPYRIGHT ALLIANCE (March 13, 2014), *available at* https://www.copyrightalliance.org/sites/default/files/statement_for_the_record_of_sandra_aistars_section_512.pdf.

³ Lisa Shuchman, “Irony 101: Citing Copyright, Sony Takes Down YouTube Video About ... Copyright,” CORPORATE COUNSEL (Feb. 17, 2016), <http://www.corpcounsel.com/id=1202749955494/Irony-101-Citing-Copyright-Sony-Takes-Down-YouTube-Video-About--Copyright> (addressing YouTube’s Content ID system and

Lumen seeks to build a corpus of real-world evidence for scholars, students, and policymakers to use to ensure that the regimes in place to police content on online platforms are properly established. Accordingly, Commenters urge the United States Copyright Office, in undertaking this study, to recognize that robust data about the sending, receipt, and efficacy of DMCA takedown notices is key to an evaluation of Section 512. Indeed, data about notices sent and received and the impact of such notices—both in specific examples and in the aggregate—is necessary in order to effectively answer most, if not all, of the questions presented in the Notice.

Commenters therefore encourage the Copyright Office to ensure that any evaluation of the Section 512 safe harbor is premised on facts and data rather than opinions and anecdotes. Commenters further ask that the Office consider taking steps to encourage those who send and receive notices to share information about those notices with the research community and the public at large.

II. PLATFORMS LIKE LUMEN ARE ESSENTIAL TOOLS FOR UNDERSTANDING THE IMPACT OF THE NOTICE-AND-TAKEDOWN REGIME.

Lumen is a research platform that invites rightsholders, Internet Service Providers and other online intermediaries, and members of the general public to share cease-and-desist letters and other forms of takedown requests that they send and receive concerning online content. Lumen maintains a database containing millions of notices that have been voluntarily shared with the project and makes those notices available to scholars, journalists, and others for purposes of research and analysis.⁴

Scholars have used this information as the basis for journal articles analyzing the consequences of policy decisions that have been made concerning the mediation of online content, including abuse of tools like the notice-and-takedown provisions of Section 512.⁵ Lumen's access to

documenting a takedown request involving teaching materials from Harvard Law School Professor and Berkman Center faculty co-director Terry Fisher).

⁴ The Lumen database may be accessed at <https://www.lumendatabase.org/>.

⁵ See, e.g., Benjamin Boroughf, *The Next Great YouTube: Improving Content ID to Foster Creativity, Cooperation, and Fair Compensation*, 25 ALB. L.J. SCI. & TECH. 95 (2015); Daniel Seng, *The State of the Discordant Union: An Empirical Analysis of DMCA Takedown Notices*, 18 VA. J.L. & TECH. 369 (2014); Leah Chan Grinvald, *Shaming Trademark Bullies*, 2011 WIS. L. REV. 625 (2011); Jennifer M. Urban & Laura Quilter, *Efficient Process or "Chilling Effects"?: Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 621 (2006); Joshua Urist, *Who's Feeling Lucky? Skewed Incentives, Lack of Transparency, and Manipulation of Google Search Results under the DMCA*, 1 BROOK. J. CORP. FIN. & COM. L. 209 (2006); Jonathon Penney, *Chilling Effects and the DMCA: An Empirical Case Study* (Mar. 5, 2016) (unpublished Ph.D. dissertation, Oxford University), available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2757778; JENNIFER M. URBAN ET AL., NOTICE AND TAKEDOWN IN EVERYDAY PRACTICE (2016), available at <http://ssrn.com/abstract=2755628>.

individual notices has also aided in popular news coverage of incidents where content has been removed online.⁶

The substance of individual notices can help flesh out narratives concerning discrete disputes over particular pieces of online content, such as when the presidential campaign of Senator Bernie Sanders sent a takedown notice to Wikimedia over campaign materials.⁷ The substance of notices in the aggregate can underscore and illuminate trends relevant to assessing the activities and motivations of those who own, upload, and host content and those who request that such content be removed.⁸

The Lumen project represents an attempt to remedy a lack of transparency about copyright takedowns inherent in Section 512, which is premised on private-party content owners sending legal notices to private-party platforms, search engines, and the like. Absent voluntary data-sharing of the sort that Lumen promotes, the Section 512 notice-and-takedown regime would operate largely without oversight or review. And, given that the Section 512 regime is a Congressional scheme for adjudicating disputes through the tailored provision of specific incentives, the act of adjudicating a Section 512 dispute has a strong public dimension.⁹ Of course, it is crucial that there be transparency in public proceedings in order to ensure public trust and legitimacy.¹⁰

III. THE QUESTIONS POSED BY THIS NOTICE OF INQUIRY ARE BEST ANSWERED BY REFERENCE TO DATA RATHER THAN ANECDOTAL EVIDENCE, AND PROJECTS LIKE LUMEN ARE THE PRIMARY MEANS THROUGH WHICH SUCH DATA IS MADE AVAILABLE.

Most of the questions posed by the current Notice of Inquiry ask, in one form or another, whether 17 U.S.C. § 512 is working properly. Commenters expect that responses to the Notice will underscore the diversity of views on the effectiveness and impact (or lack thereof) of the

⁶ See, e.g., Abby Ohlheiser, "The Internet's most famous dog rater keeps disappearing from Twitter," THE WASHINGTON POST (Feb. 11, 2016), <https://www.washingtonpost.com/news/the-intersect/wp/2016/02/11/the-internets-top-dog-rater-keeps-disappearing-from-twitter/> (citing notices in Lumen's database in reporting on story about online dog ratings).

⁷ See Joe Mullin, "Bernie Sanders lawyers to Wikipedia: Take down our logo, you're violating DMCA [Updated]," ARSTECHNICA (January 15, 2015), <https://www.arstechnica.com/tech-policy/2016/01/bernie-sanders-campaign-sends-dmca-notice-to-wikipedia-over-logos/> (linking to and discussing Lumen's hosting of that message at <https://www.lumendatabase.org/notices/11716729>).

⁸ See Seng, *supra* note 5; Caitlin Dewey, "How we're unwittingly letting robots censor the Web," THE WASHINGTON POST (March 29, 2016), <https://www.washingtonpost.com/news/the-intersect/wp/2016/03/29/how-were-unwittingly-letting-robots-censor-the-web/> (discussing the value of data in analyzing trends in automation of content removal).

⁹ See generally Rebecca Tushnet, *Power Without Responsibility: Intermediaries and the First Amendment*, 76 GEO. WASH. L. REV. 986 (2008) (extensively analyzing the relationship between government and private power around the DMCA notice-and-takedown regime).

¹⁰ See *Press-Enterprise Co. v. Superior Court*, 464 U.S. 501, 508 (1984) (noting in the context of access to trials that openness "enhances both the basic fairness . . . and the appearance of fairness so essential to public confidence in the system").

current DMCA notice-and-takedown regime. From characterizing the DMCA safe harbor as “the law that saved the web,”¹¹ to a “loophole” via which “artist rights get destroyed,”¹² commentators have no shortage of opinions.

Notably, facts are often missing from these kinds of debates.¹³ Data-driven or evidence-based approaches to policymaking have been advanced, embraced, and implemented effectively in various areas of law and industry to great effect.¹⁴ Indeed, it is axiomatic that data-driven policy decisions are preferred, as they are “more precise, because they [are] more informed.”¹⁵ There is no reason not to apply this approach with respect to regulation of the Internet.

As Internet scholars John Palfrey and Jonathan Zittrain noted in an article in *Science* magazine, decisions regarding Internet regulation have particularly far-reaching implications given such decisions’ “profound societal impact” and connections to “economics, free expression and privacy.”¹⁶ Expressing concerns about the “paucity of data available to guide” policy about online activities, they concluded:

The best approach is neither to make ill-informed decisions based on too little data nor to avoid state regulation simply because of the absence of decent data. Instead, we should begin a concerted push for highly reliable and publicly available forms of measurement of the Internet and the Web and how we use them, including the flows of information we generate and consume. Better data would do more than just help the state meet its regulatory obligations; better data would also improve self-regulation by private sector players and empower individuals to make better decisions. In the meantime, we as researchers need to work harder to

¹¹ David Kravets, “10 Years Later, Misunderstood DMCA is the Law That Saved the Web,” WIRED (Oct. 27, 2008), <http://www.wired.com/2008/10/ten-years-later/>.

¹² Take Down Stay Down Petition, TAKE DOWN STAY DOWN, <http://www.takedownstaydown.org/> (last visited Mar. 30, 2016).

¹³ See generally WILLIAM PATRY, MORAL PANICS AND THE COPYRIGHT WARS (2009) (extensively analyzing the absence of facts in copyright debates).

¹⁴ *Legislating Evidence-Based Policymaking*, THE PEW CHARITABLE TRUSTS, MACARTHUR FOUNDATION (March 2015), <http://www.pewtrusts.org/~media/assets/2015/03/legislationresultsfirstbriefmarch2015.pdf?la=en> (discussing general data-driven or “evidence-based” policymaking and providing examples of such practices). See also, Adam Gelb, “State Criminal Justice Reforms Build the Case for Data-Driven Federal Legislation,” THE PEW CHARITABLE TRUSTS, MACARTHUR FOUNDATION (July 30, 2015), <http://www.pewtrusts.org/en/research-and-analysis/analysis/2015/07/30/state-criminal-justice-reforms-build-the-case-for-data-driven-federal-legislation>; Madelyn L. Isaacs, *Data-Driven Decision Making: The Engine of Accountability*, 6 PROF. SCH. COUNSELING 288, 288 (2003) (discussing the role of data-driven policy in education).

¹⁵ Lauren Alexander, “3 Reasons Why The Future of Policy Making Will be Data-Driven,” GRANICUS (Dec. 17, 2014), <http://www.granicus.com/civic-technology/3-reasons-why-the-future-of-policy-making-will-be-data-driven/>.

¹⁶ John Palfrey & Jonathan Zittrain, *Better Data for a Better Internet*, SCIENCE, Vol. 334, Issue 6060 (December 2, 2011), available at <http://science.sciencemag.org/content/334/6060/1210.full?ijkey=yLsWDbb0ekI&keytype=ref&siteid=sci>.

translate the good data that we do have into terms that can directly inform policy-making.¹⁷

A comprehensive evaluation of Section 512 requires an assessment of how the provision strikes a balance between legitimate aims of content owners (who seek to protect their intellectual property) and the strong interest in free expression enshrined in the First Amendment to the United States Constitution (which is hostile to government-endorsed efforts to restrain speech absent due process and strict tailoring). In undertaking such an evaluation, one must look to the effects of notices sent and received. By working with private parties that engage in the process of sending and responding to takedown notices, the Lumen project has created a repository of precisely the kind of data needed to develop and hone sound Internet policy in this space.

For example, the following categories of data available in the Lumen database might be relevant in answering the various questions that comprise this Notice:

- ***Data regarding the overall volume of notices sent and received over time.*** Since its inception, Lumen has observed a dramatic increase in the number of notices it receives, from an average of one (or fewer) per day in 2002, to between four- and five-thousand per day this year. It is relevant to this Office to consider the nature of how that dramatic increase took place. The Office may be surprised to learn, for example, that the increase is more exponential than linear; it took over ten years for Lumen to receive its one-millionth notice but only a little over a year to receive its two-millionth, less than one year for the third million, and only eight months for the fourth million. When compared to the overall rise of user-generated content,¹⁸ the similarities and differences between the increase of content and the increase of DMCA actions provide important insights into the nature of rights enforcement online.
- ***Data relevant to trends in numbers of discrete URLs embodied in each notice.*** While much public discussion focuses on the number of notices sent or received, it is crucial that this Office understand that this metric obscures the actual volume of content being taken down. When Lumen began to collect notices, each DMCA notice typically contained only one or two URLs, usually accompanied by a personalized letter.¹⁹ More recent notices often contain many more URLs, sometimes one-thousand or more in a single notice.²⁰

¹⁷ *Id.*

¹⁸ See, e.g., Pamela J. McKenzie et al., *User-Generated Online Content: Overview, Current State, and Context*, FIRST MONDAY (June 2012), <http://firstmonday.org/article/view/3912/3266>.

¹⁹ See DMCA Takedown Notice Regarding Posting of Scientology Excerpts, LUMEN (Feb. 8, 2002), <https://www.lumendatabase.org/notices/122>.

²⁰ See DMCA Copyright Complaint to Google from BPI, LUMEN (Mar. 30, 2016), <https://www.lumendatabase.org/notices/12012938>. BPI, the sender of this notice (and one of the largest senders by volume in Lumen's database) has requested the removal of over 200 million URLs from Google alone, at an average of 607 URLs per notice. While this number itself is noteworthy, BPI used 274,810 DMCA notices to request the

- ***Data relevant to trends in content owners’ reliance on outside vendors to facilitate the sending of takedowns.*** The Notice addresses concerns about how the DMCA may be affecting different players in the space differently. Among other data points, Lumen seeks to track the identities of senders and recipients of notices in its database, along with the “principal” or ultimate complaining rightsholder on whose behalf a given notice is sent. It is therefore relevant to this inquiry that of the top ten senders currently represented in the more than four-million notices in Lumen’s database, seven are third-party outside vendors or rights management companies. The most prolific sender, Audiolock.net, has sent nearly one-million notices.
- ***Data relevant to trends in automation of the sending and receipt of DMCA notices.*** A recent study that builds on data drawn from Lumen suggests that, although the vast majority of websites use humans to process notices sent and received, among social networks, search engines and file hosts—the sites that get the most complaints—it is increasingly common to use automated triage-and-escalation systems similar to those used by third-party vendors.²¹

In addition to the work of Lumen, some websites have started publishing their own transparency reports that collect, analyze and publicize general data about their receipt of and the consequences flowing from takedown notices of various kinds (including demands sent by both government and private parties). Some of these reports provide specific information about the types of requests received and any actions taken in response. The value of such transparency reporting must not be understated.²² Nevertheless, while transparency reporting represents a growing trend among some of the larger online services, many sites do not undertake this type of reporting. Those that publish transparency reports do not generally include the text of notices they receive. And, specific information on a site-by-site basis not inform the greater picture of the state of DMCA notices to the same extent as aggregated information.²³ The work of projects like Lumen therefore represent not a replacement for transparency reporting but, rather, an important analog thereto.

removal of their first one hundred million URLs (an average of ~384 URLs per notice). It subsequently used fewer than 55,000 notices to request removal of the next 100 million URLs, an average of over 1,800 URLs per notice.

²¹ URBAN ET AL., NOTICE AND TAKEDOWN, *supra* note 5.

²² LIZ WOOLERY ET AL., THE TRANSPARENCY REPORTING TOOLKIT: BEST PRACTICES FOR REPORTING ON U.S. GOVERNMENT REQUESTS FOR USER INFORMATION (2016), available at https://cyber.law.harvard.edu/publications/2016/transparency_memos.

²³ See, e.g., *Transparency Report: Copyright notices*, TWITTER, <https://transparency.twitter.com/copyright-notices/2015/jul-dec> (last visited Mar. 31, 2016); *Transparency Report: Requests for Content Alteration & Takedown*, WIKIMEDIA FOUNDATION <https://transparency.wikimedia.org/content.html> (last visited Mar. 31, 2016); *Transparency Report*, GOOGLE, <https://www.google.com/transparencyreport/> (last visited Mar. 31, 2016); *Intellectual Property*, AUTOMATTIC, <https://transparency.automattic.com/intellectual-property/> (last visited Mar. 31, 2016); *Kickstarter Transparency Report 2014*, Kickstarter (Apr. 8, 2015), <https://www.kickstarter.com/blog/kickstarter-transparency-report-2014> (providing examples of site-specific “transparency reports”).

For all of these reasons, Commenters urge the Copyright Office to embrace a data-driven approach to its assessment of Section 512 and to encourage those who send and receive notices to share those notices with researchers and the general public.

IV. THE COPYRIGHT OFFICE SHOULD TAKE STEPS TO ENCOURAGE TRANSPARENCY AROUND TAKEDOWNS.

Transparent processes around the DMCA do not just help inform general policy. They also ensure that online actors across the spectrum are held accountable for their actions and that individuals affected by DMCA takedowns are afforded due process. Absent independent organizations gathering and sharing this information, the process of requesting content removal under Section 512 and the process of effectuating such requests take place outside the public eye. Making notices available permits the researchers, journalists, and the general public to monitor the takedown system, both as it applies to individual cases and in the aggregate.²⁴ Enabling and empowering monitoring increases the chances of identifying problems (systemic or party-specific) in a regime that is so immense as to render internal policing by individuals impacted by takedown notices practically impossible.²⁵

It is worth noting in this regard that Section 512 provides for removal under circumstances that considers the rights of those who upload content to a far lesser degree than US law typically considers the rights of speakers, offline and online.²⁶ In other contexts, scholars and courts have noted the tension between the availability of preliminary relief in cases involving intellectual property claims and courts' traditional abhorrence of prior restraints on speech, questioning whether copyright injunctions unconstitutionally restrict speech.²⁷

Such commentators take for granted that, whatever their flaws, restraints in copyright cases are judicially imposed and come with the opportunity for transparency and oversight inherent in judicial proceedings. Procedures governing content removal under the DMCA, on the other hand, take place almost entirely outside of the traditional judicial system and thus raise the specter of

²⁴ See "Music Group DMCA Notices Reveal Coffee Hatred," TORRENTFREAK (Aug. 25, 2014), <https://torrentfreak.com/music-group-dmca-notices-reveal-coffee-hatred-140825/> (identifying a potential abuser of the DMCA in reliance on Lumen's data).

²⁵ See Chris Welch, "Google received over 75 million copyright takedown requests in February," THE VERGE (Mar. 7, 2016), <http://www.theverge.com/2016/3/7/11172516/google-takedown-requests-75-million>.

²⁶ Indeed, no court has fully addressed the question of the system's constitutionality. See, e.g., Tushnet, *supra* note 9, at 1004–05.

²⁷ See Mark A. Lemley & Eugene Volokh, *Freedom of Speech and Injunctions in Intellectual Property Cases*, 48 DUKE L.J. 147 (1998). After asserting that copyright law restricts speech, Lemley and Volokh argue that injunctions in copyright cases, especially preliminary injunctions, might violate the constitution because of the lack of necessary prior procedure (i.e., a final judicial determination). *Id.* at 165 ("Copyright law restricts speech: it restricts you from writing, painting, publicly performing, or otherwise communicating what you please."); 169–171, 174 (noting that prior restraints on speech are almost always treated as unconstitutional prior to a final judicial determination). See also *Salinger v. Colting*, 607 F.3d 68, 82 (2d Cir. 2010) ("Every injunction issued before a final adjudication on the merits risks enjoining speech protected by the First Amendment.").

abuse or overbroad application. Indeed, in the modern era of the DMCA, some have argued that Section 512 directly incentivizes intermediary entities “to take down content that is protected by fair use and the First Amendment.”²⁸

Against this backdrop, the need for oversight, evaluation, and review of processes governing content removal is essential. With increased transparency—and a resultant increase in accountability and oversight—parties can use the system to more effectively advocate for their own rights in the face of claims that interfere with freedom of expression.

V. CONCLUSION

As set forth herein, Commenters urge the Copyright Office to look to data to inform its evaluation of Section 512 and to keep in mind the important role such data plays in setting and enforcing policies concerning online speech and copyright enforcement. Commenters further ask that the Office consider ways to encourage information-sharing among parties that send and receive takedowns, the research community, and the public at large. Only by collecting, reviewing, and evaluating requests for content removal—individually and *en masse*—can we simultaneously develop robust and effective policy and facilitate the level of scrutiny and oversight that ought to characterize any regulatory framework that impacts the ability of Internet users to engage in free expression.

Respectfully submitted,



Christopher T. Bavitz
Managing Director, Cyberlaw Clinic²⁹
Clinical Professor of Law, Harvard Law School
1585 Massachusetts Avenue
Suite 5018
Cambridge, MA 02138
Tel: 617-384-9125
Email: cbavitz@cyber.law.harvard.edu

on behalf of
Adam Holland and Christopher T. Bavitz

²⁸ Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2314 (2014).

²⁹ Commenters thank spring 2016 Cyberlaw Clinic students Jonathan Luebbers and Shoshana Schoenfeld of Harvard Law School for their valuable and significant contributions to this comment.